Occurrence of Exceedances in a Finite Perpetuity

A thesis submitted to the Faculty of Mathematical Sciences
at the University of Oxford in Trinity term 2004
by Nathanaël Alexandre BENJAMIN of St Peter's College
for the title of Doctor of Philosophy.

Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, United Kingdom

Abstract

Generated by stochastic recursions, perpetuities encompass a vast range of discrete-time financial behaviours. When focusing on the dramatic changes occurring in such processes, the analysis of threshold exceedances provides an extensive description of their underlying mechanisms. Asymptotically, an exceedance point process tends to a compound Poisson measure, highlighting a tendency to cluster. Now, the parameters of this limit law are known, but complex. Here, an empirical approach is adopted, and a class of explicit compound Poisson models developed, with a bound on the error, for the exceedance point process of a finite, multidimensional perpetuity. In a financial regulatory context, this provides a new way of examining the Value-at-Risk criterion for securities.

Acknowledgements

It is a difficult exercise to put certain things into words. The years I spent writing this thesis have been intensely fulfilling.

I was very lucky to work under the supervision of Dr Gesine Reinert. It has been a real delight, and her kindness, availability, acuteness and enthusiasm were powerful sources of inspiration. My gratitude also goes to Professors Thomas Mikosch and Sir David Cox, for some extremely insightful comments at various stages of my research, to Carla Ysusi for key programming hints and to Alastair Pickett for useful LaTeXtricks.

The unwavering support of my family has been a constant source of serenity, and the fondness with which I look back on my time in Oxford is also due to the brilliant friends I was fortunate enough to make there, both in St Peter's College and in the Department of Statistics.

Contents

Chapter 1

Introduction

Accelerated occurrence of catastrophes has been a marking feature of the past three decades. Reasons for this phenomenon are multiple. Climate change might be at the origin of unusually massive rainfall, with consequences like the recent 2002 floods in Central Europe and the 1999 landslides in Venezuela. The intensification of traffic around the globe, on the seas and in the air, went along with an increase of accidents, air crashes or shipwrecks. Simultaneously, the constant swelling of financial market volumes leads to rocketing volatilities.

Because of this, the insurance sector is faced with new challenges, putting the relevance of old actuarial techniques to the test. In the last few decades, probabilists have developed extreme value theory to provide a new class of tools.

One of the basic rules of risk policies in the banking sector is that the more accurately a bank assesses its financial risk, the more generous the financial authorities are, which means that they allow the bank to take greater risk, for instance by

lowering its minimum capital requirements. The original object of the 1988 Basel Accord [9], also called "Basel I", developed by the Basel Committee on Banking Supervision of the Bank for International Settlements (BIS), and endorsed by the central bank governors and heads of bank supervisory authorities of the Group of Ten (G-10) countries, was credit risk, essentially the risk that the financial situation of a debitor might worsen, or that the debitor might even default. The 1996 Amendment [10], also called "BIS 98" after it was implemented, focused essentially on the market risk arising from the trading book. Risk is currently to be measured by the value-at-risk (VaR), as defined in the Basel Capital Accord 1988, and later in the very recent 2004 Revised International Capital Framework [11], also called "Basel II". The VaR is three times the estimated 1%-quantile of the profit/loss distribution for the next ten days based on observations of at least one year. Yet, the use of VaR as a measure of risk is controversial, for instance because it is not subadditive, which means that there are cases when a portfolio diversification, even though safer, can lead to a larger VaR figure; *c.f.* Daníelsson *et al.* [23] for a broader discussion of VaR as a risk measure.

The quest for accuracy implies a good estimation power, which is achieved by using past data. The problem is that, for instance in the case of operational risk, addressed in Basel II, which accounts for fraud or potential mishandlings when transactions are implemented, many banks did not keep records of relevant information. Hence, they now have to simulate such data. The available models for extremal events are provided by extreme value theory, but this does not solve everything. First, a safe risk measure should take the modelling error into account. Second, models often describe the asymptotic behaviour. But life is not asymptotic. One only ever has a finite number of data to deal with. The following work is based on these two ideas.

In addition, in the more standard case of market risk, the object of the 1996 Amendment, assessing potential portfolio losses following changes in market levels, a number of problems related to modelling arise. Essentially, there are three main methods to calculate market VaR, *c.f.* Crouhy *et al.* [22]. All of them try to evaluate

the profit and loss distribution of the portfolio over the period considered. The first method is the variance-covariance approach. It assumes that the components of the portfolio are multivariate normal. The portfolio, being a linear combination of them, is also normal. The parameters of the multivariate normal distribution are estimated from past data. Combined with the component weights and correlations, they give the parameters of the portfolio distribution, and hence the portfolio VaR. The second method is the historical approach, where the quantile is directly estimated from the empirical portfolio distribution. The third classic method is the Monte-Carlo approach. Again, multivariate normality is assumed and the relevant parameters estimated from the data. Then portfolio configurations are simulated. One value is simulated for the first asset, then, given that value and the covariance matrix, a value is simulated for the second asset, and so on, until the whole portfolio has been described. This gives one portfolio value. Many others are then simulated, and a distribution curve is obtained. Finally, the VaR is estimated form this curve. The obvious problem with the second method is that the VaR estimator is clearly not robust. The essential issue with the first and third calculation methods is that they rely on normal assumptions. Normal distributions are designed to represent usual events and discard unusual ones. Now VaR tells us about the occurrence of unusual events, therefore the use of normal distributions to describe them is questionable. Indeed, the estimation of normal variables is less reliable in the tails, and this is precisely what we are interested in. This is why extreme value theory, and more precisely the analysis of threshold exceedances, offers an opportunity to calculate VaR with proper tools. An application of extreme value theory to Value-at-Risk can be found in Borkovec and Klüppelberg [17], and also in Longin [50], were an approach similar to the variance-covariance one is taken for the study of multiple risk factors, even though the VaR of each covariate is computed with extreme distributions instead of Gaussian ones. Let us now say a word on the sort of processes we shall be investigating.

Perpetuities, *i.e.* the random variables generated by stochastic recursions, are widely used in finance to describe the behaviours of stock returns, exchange rates, futures, interest rates, and other market-driven entities. The asymptotic theory for the tail of variables verifying a stochastic recurrence equation has been first explored by Kesten [44], Vervaat [67] and Goldie [36], and a thorough insight on extreme value theory can be found in Leadbetter *et al.* [47] and Embrechts *et al.* [30].

An essential result in this field is the existence of a sequence of threshold values, such that the number of exceedances above them has an asymptotic compound Poisson distribution, *c.f.* Hsing *et al.* [41]. Compound Poisson approximations are widely used in the modelling of extremes for insurance and finance, and an asymptotic theory for them can be found *e.g.* in Barbour and Chryssaphinou [5]; let us be a bit more specific. For an i.i.d. stochastic process, the asymptotic distribution of the number of exceedances above the relevant sequence of thresholds is a homogeneous Poisson distribution. Now, if one allows for time correlation in the process, the exceedances will have a tendency to occur in clusters. For instance, looking at hail risk, *c.f.* Klüppelberg [45], whenever this happens, the odds for storms and floods to happen as well are greater, and these will generate yet another set of large, numerous claims. The claims for hail and those for the floods constitute a cluster of exceedances, and the time correlation corresponds to the fact that similar causes generate hail and floods. In these cases of time correlation, the asymptotic distribution will switch from a homogeneous Poisson to a compound Poisson distribution, composed of the homogeneous Poisson distribution of the clusters, mixed with the distribution of the cluster sizes, which are usually assumed to be i.i.d. An underlying assumption is stationarity. This might be reasonable in a financial framework, if the data considered is taken between two structural breaks, even though these are difficult to identify. However, in an environmental setting, climate change issues might make such an assumption hazardous.

Whether through sequential maxima or threshold exceedances, the univariate

theory of extremes has now been thoroughly explored, and many issues have been successfully addressed, *c.f.* for instance Embrechts *et al.* [30]. The biggest challenge now facing extreme value theory reseachers is the multidimensional setting. Knowing the marginal extremal behaviour of covariates does not give us a full description of the whole vector, because it does not answer the question of dependence between covariates. How does a risk factor react when another one has gone beyond a given threshold? This problem is being intensely examined by the scientific community at the moment. One powerful notion for this purpose is the one of regular variation, *c.f.* Bingham *et al.* [15], where a spectral measure describes the dependence between covariates. Recent work on multivariate extremes can be found in Capéraà *et al.* [20] and Heffernan and Tawn [40] in a general setting, Starica [64] in a financial context, and Embrechts *et al.* [31] with a copula approach.

A model can be more or less appropriate, and when it is involved in a risk measurement, the need for a bound on its accuracy is crucial. A Wasserstein bound on the Poisson approximation to a Bernoulli process, derived using Stein's method, can be found in Xia [68] and Barbour, Holst and Janson [6]. Now, the argument against the homogeneous Poisson approach is that one would have to assume time independence in order to obtain a Poisson process of exceedances. Overcoming this obstacle, total variation bounds on the compound Poisson approximation have been found, first in Arratia, Gordon and Goldstein [2], then in Barbour, Chen, and Loh [4], and also later in Barbour and Utev [8]. An overview of total variation results for the compound Poisson approximation with Stein's method can be found in Barbour and Chryssaphinou [5]. Kerstan's method is used for the same purpose in Roos [62]. Now, these are theoretical results, which are not optimal for the analysis of exceedances. The argument against the total variation analysis is that this metric is not fit to describe distances between continuous and discrete random measures. Taking into account the various contingencies mentioned above, a theoretical Wasserstein bound on the compound Poisson approximation for the distribution of extremes has

been recently found by Barbour, Novak and Xia [7]. Because analysts are mostly dealing with finite samples, the empirical approach adopted by the authors, which does not require any knowledge of parameters for the asymptotic behaviour of the process, is highly valuable. The object to approximate is the empirical point process of exceedances, for a finite sequence of events. The coefficients of the approximating compound Poisson measure are derived from the original process. Under suitable convergence conditions, they match the asymptotic results, *c.f.* O'Brien [56].

The work presented in this thesis was originally motivated by a study of the dramatic fluctuations of Dollar-Sterling exchange rates between 1992 and 2001, in Benjamin [13]. Exchange rates often show some conditionally heteroscedastic characteristics, which make GARCH models particularly popular for them, *c.f.* Bollerslev [16]. The squared volatility of GARCH(1,1) processes follows a stochastic recurrence equation of the first order, and hence forms a perpetuity, to which theoretical results can therefore be applied.

The aim of this work is to link the theory of empirical point processes to the one of exceedances in stochastic recursions. The approximation found in Barbour *et al.* [7] is improved for this special framework, and once put together with a judicious m-dependent approximation, leads to an overall compound Poisson model. In an ultimate stage, the approximation is performed in the more general framework of hypercube exceedances of multidimensional perpetuities, which constitutes the essential contribution of this work to the current reflexions on multivariate extreme value theory. Finally, the compound Poisson measure obtained is used to simulate the occurence of exceedances, and obtain an estimation of the VaR.

Now, more precisely, a one-dimensional perpetuity is a random variable on the state space $\mathbb{R}$ generated by a stochastic recursion of the form

$$X_t = B_t X_{t-1} + A_t, \tag{1.1}$$

where $\big((A_t, B_t)\big)_{t\in\mathbb{N}^*}$ are independently and identically jointly distributed on $\mathbb{R}^2$. We

shall consider a strictly stationary finite sequence of size n generated by (1.1). Conditions for the existence of such a stationary solution will be presented in Chapter 6. We shall then focus on its empirical point process of exceedances above a given threshold $u \in \mathbb{R}$, *i.e.*

$$N_n(\bullet) \;=\; \sum_{t=1}^{n} \mathbb{I}_{\{X_t > u\}} \delta_{tp}(\bullet),$$

where $p = \mathbb{P}(X_1 > u)$. This point process N_n is a random measure on the measurable space $[0,1]$. The aim of this work is to approximate it by an appropriate compound Poisson process. In the multidimensional setting, we shall look at random recursions on the state space $\mathbb{R}_+^d$, $d \in \mathbb{N}^*$, of the form

$$\boldsymbol{X}_t = \boldsymbol{B}_t \boldsymbol{X}_{t-1} + \boldsymbol{A}_t, \tag{1.2}$$

where $\big((\boldsymbol{A}_t, \boldsymbol{B}_t)\big)_{t\in\mathbb{N}^*}$ are independently and identically jointly distributed. The random vectors $(\boldsymbol{A}_t)_{t\in\mathbb{N}^*}$ take their values in $\mathbb{R}_+^d$, and the random matrices $(\boldsymbol{B}_t)_{t\in\mathbb{N}^*}$ are of size $d\times d$, with non-negative entries. Conditions for the existence of a stationary solution are provided in Chapter 8. In this new framework, we shall look at exceedances outside d-dimensional parallelepipeds of the type $R_{0,\boldsymbol{u}} = [0,u_1]\times\cdots\times[0,u_d]$, and the related empirical point process of exceedances is

$$N_n(\bullet) \;=\; \sum_{t=1}^{n} \mathbb{I}_{\{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}} \delta_{tp}(\bullet),$$

where $p = \mathbb{P}(\boldsymbol{X}_1 \notin R_{0,\boldsymbol{u}})$. Again, N_n is a point process on $[0,1]$, which we shall approximate by a compound Poisson measure.

In Chapter 2 we recall some useful results on point processes, Markov chains and mixing processes, then Chapter 3 deals with the extreme value theory background upon which this thesis is built. In Chapter 4, we give a heuristic illustration of the overall compound Poisson approximation in a practical case of exchange rate volatilities. Now, a stochastic process of exceedance indicators is $\{0,1\}$-valued, and this is what we want to model. In a first stage, we shall ignore any underlying process

of which it might be a hidden Markov chain, and directly study a few $\{0,1\}$-valued processes. This is why we adapt the compound Poisson approximation presented in Barbour *et al.* [7] in the simple case of a $\{0,1\}$-valued first order Markov chain in Chapter 4, and the case of a d-th order chain is developed in Chapter 5. In Chapter 6, we take into consideration the fact that the $\{0,1\}$-valued process of interest is generated by a perpetuity, and use the recursive structure of the latter to approximate it by an m-dependent process. We generalise this approach to the multidimensional case in Chapter 7, and make the final step in Chapter 8, where the $\{0,1\}$-valued m-dependent processes obtained are linked to a compound Poisson measure. Chapter 9 sums up the whole approximation results, and the results are illustrated in Chapter 10, with an application to the evaluation of the Value-at-Risk criterion, and Chapter 11 is the conclusion of the thesis.

Chapter 2

Point processes and stochastic stability

Point processes are not just a count of events, but also contain all the information on their locations in time or space. Their dual nature, both probabilistic and topological, makes them powerful, and therefore they are very useful mathematical tools to describe behaviours and features.

2.1 Discrete random measures

In a first stage, we need to introduce two fundamental notions, following Kallenberg [43], and leading to that of a point process. Recall first that a measure on a given measurable space is said to be *locally finite* if any point of the space has a neighbourhood where the measure is bounded.

Definition 2.1.1 (Random measure) *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space. Let* $(S, \mathcal{S})$ *be a measurable space. A* random measure *is a mapping* ξ *from* $\Omega \times \mathcal{S}$ *to* $[0, +\infty]$, *such that for all* ω *in* Ω, $\xi(\omega, \bullet)$ *is a locally finite measure on* S, *and for all* A *in* $\mathcal{S}$, $\omega \mapsto \xi(\omega, A)$ *is a random variable.*

We shall now suppose that S is a Polish space and that $\mathcal{S}$ is the Borel σ-field of S, denoted by $\mathcal{B}(S)$, which makes $(S, \mathcal{S})$ a Borel space. The class of measures on $(S, \mathcal{B}(S))$ we shall consider in the future is that of the σ-finite measures on S, which set we shall denote by $\mathcal{M}(S)$. Discrete and continuous measures can be used on S, and counting measures are an example of the former type. We refer to the definition in Barbour *et al.* [7].

Definition 2.1.2 (Counting measure) *Given the following notation for the Dirac measure on* S, *where* $\mathcal{S}(S)$ *is the set of subsets of* S,

$$\forall M \in \mathcal{S}(S), \forall s \in S, \quad \delta_s(M) = \begin{cases} 1 & \text{if } s \in M \\ 0 & \text{if } s \notin M \end{cases}$$

a counting measure *on* S *associated with a sequence of elements* $(s_i)_{i \in \mathbb{N}^*}$ *of* S *is an element* $N(\bullet)$ *of* $\mathcal{M}(S)$ *of the form*

$$N(\bullet) = \sum_{i=1}^{+\infty} m_i \delta_{s_i}(\bullet)$$

where

$$\forall i \in \mathbb{N}^*, \quad m_i \in \mathbb{N}.$$

Denoting by $\mathcal{N}(S)$ the set of counting measures on S, we are now in a position to finally give a definition for point processes, as in Kallenberg [43], and then interpreting it as a counting measure with random masses.

Definition 2.1.3 (Point process) *A* point process *on* S *is an integer-valued random measure on* S. *Restricting ourselves to non-negative values, it is a random*

measure of the form

$$\begin{aligned} \Omega &\longrightarrow \mathcal{N}(S) \\ \omega &\longmapsto N(\omega, \bullet) = \sum_{i=1}^{+\infty} m_i(\omega)\delta_{s_i}(\bullet) \end{aligned}$$

where

$$\forall i \in \mathbb{N}^*, \forall \omega \in \Omega, \quad m_i(\omega) \in \mathbb{N}, \quad s_i \in S\,.$$

When the sequence (s_i) of elements in S is finite, the corresponding point processes are called *empirical point processes.* They are of the form

$$N_n(\omega, \bullet) = \sum_{i=1}^{n} m_i(\omega)\delta_{s_i}(\bullet)\,.$$

Finally, a point process is said to be *simple* if its random weights $m_i(\omega)$ all take values in $\{0, 1\}$.

Example 2.1.1 (Empirical exceedance point process) Let $n \in \mathbb{N}^*$. Let $(X_i)_{1 \le i \le n}$ be a sequence of random variables with values in $\mathbb{R}$. Let $u \in \mathbb{R}$. Choosing $S =]0, 1]$ and $\forall i \in [\![1, n]\!]$, $s_i = i/n$, we can form the following empirical point process,

$$N_n(\bullet) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i \ge u\}}\delta_{i/n}(\bullet)$$

called *empirical point process of exceedances* of the threshold u within the sequence $(X_i)_{1 \le i \le n}$.

Remark. The space classically chosen for exceedance occurrences is $]0, 1]$, with the related appropriate scaling in the indices of the Dirac measures, because these empirical processes will be approximated by continuous random measures. Thus restricting the space to $]0, 1]$ ensures that the limiting process also takes values in $]0, 1]$ only.

The considerable advantage of an empirical point process of exceedances is that it not only counts threshold exceedances within a sequence, but tells us as well where

they occur. This is most precious when we are only given a finite time sequence, because we are still able to describe the evolution in time of its exceedance behaviour, all this information being contained in just one variable.

2.2 Compound Poisson processes

Empirical point processes count variables in a space, and the ones we shall deal with have an asymptotic compound Poisson behaviour. Let us define what we mean by a *Poisson process* or *Poisson random measure.* More can be found about these measures in Barbour *et al.* [6]. A general definition of a Poisson process is given as in Kallenberg [43], and two classes of such processes are then presented.

Definition 2.2.1 (Poisson point process) *A point process is said to have* independent increments *if its values on two disjoint relatively compact elements of $\mathcal{B}$ are independent random variables. A* Poisson process *or* Poisson random measure *on S with intensity measure $\mu \in \mathcal{M}(S)$ is a point process with independent increments such that its value on any relatively compact element B of $\mathcal{S}$ has a Poisson distribution with mean $\mu(B)$.*

Now, with the same notations as in the previous section, a discrete *Poisson random measure carried by fixed points $(s_i)_{i\in\mathbb{N}^*}$ in the space S with intensity $\lambda \in \mathbb{R}$ is a point process of the form*

$$N(\omega, \bullet) = \sum_{i=1}^{+\infty} m_i(\omega)\delta_{s_i}(\bullet) = \text{Poisson}\big(\lambda \sum_{i=1}^{+\infty} \delta_{s_i}(\bullet)\big)$$

where

$$\forall\,(i,k) \in \mathbb{N}^* \times \mathbb{N}, \forall\,\omega \in \Omega, \quad m_i(\omega) \in \mathbb{N}, \quad \mathbb{P}\,(m_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}\ .$$

When the state space is a fixed interval $[0, b]$, with $b \in \mathbb{R}^{+*}$, a continuous *Poisson random measure on $]0, b]$ with intensity $\lambda\boldsymbol{\mu}(\bullet)$, where $\boldsymbol{\mu}(\bullet)$ is the Lebesgue measure*

on $[0,b]$, *is a point process of the form*

$$N(\omega,\bullet)=\sum_{i:b_i\in[0,b]}\delta_{b_i}(\bullet)=\text{Poisson}\big(\lambda\boldsymbol{\mu}(\bullet)\big)$$

where $(b_i)_{i\in\mathbb{N}^*}$ *are random points of* S *such that*

$$\forall\,(i,k)\in\mathbb{N}^*\times\mathbb{N},\quad \mathbb{P}\left(\text{card}\{i:b_i\in[0,b]\cap\bullet\}=k\right)=\frac{(\lambda\boldsymbol{\mu}(\bullet))^k e^{-\lambda\boldsymbol{\mu}(\bullet)}}{k!}\,.$$

The actual approximating processes we shall be dealing with will in fact be a special sort of inhomogeneous Poisson random measure: the compound Poisson measures, where homogeneous Poisson processes of clusters are weighted by the i.i.d. cluster sizes. Recall that the compound Poisson distribution with intensity λ and compounding distribution ν is the law of a random variable of the form $\sum_{i=1}^{N} Z_i$, where N is a Poisson(λ) random variable independent of N, and (Z_i) are i.i.d. random variables with distribution ν.

Definition 2.2.2 (Compound Poisson process) *With the notations of the previous definition, a discrete* compound Poisson random measure *or* compound Poisson process *on a space* S *with intensity* $\lambda\in\mathbb{R}^+$ *and compounding distribution* ν *is of the form*

$$N(\omega,\bullet)=\sum_{i=1}^{+\infty}Z_i(\omega)m_i(\omega)\delta_{s_i}(\bullet)=\text{CP}\big(\lambda\sum_{i=1}^{+\infty}\delta_{s_i}(\bullet),\nu\big)$$

where $\{Z_i\}$ *is independent of* $\{m_i\}$, *and*

$$\forall\, i\in\mathbb{N}^*,\forall\,\omega\in\Omega,\quad Z_i(\omega)\in\mathbb{N},\quad Z_i\overset{d}{=}\nu\,,$$

and also, as in the previous definition,

$$\sum_{i=1}^{+\infty}m_i(\omega)\delta_{s_i}(\bullet)\;=\;\text{Poisson}\big(\lambda\sum_{i=1}^{+\infty}\delta_{s_i}(\bullet)\big).$$

Now, in the continuous intensity measure framework, a compound Poisson random measure with intensity $\lambda\boldsymbol{\mu}(\bullet)$ and compounding distribution ν is a point process of the form

$$N(\omega,\bullet)=\sum_{i:b_i\in[0,b]}Z_i(\omega)\delta_{b_i}(\bullet)=CP\big(\lambda\boldsymbol{\mu}(\bullet),\nu\big)$$

with the same definition as above for the cluster sizes $(Z_i)_{i\in\mathbb{N}^*}$.

2.3 Metrics for point processes

One of the aims of this thesis is to approximate some empirical point processes by compound Poisson processes. We need to specify which metrics we shall use to calculate the distance between model and reality.

Definition 2.3.1 (Total variation distance) *Let $\mathcal{P}(S)$ be the set of point processes on S. Let $\mathcal{N}(S)$ be the set of counting measures on S. Let $\mathcal{B}\circ\mathcal{N}(S)$ be the Borel set of $\mathcal{N}(S)$. Let η and ξ be distributions of elements of $\mathcal{P}(S)$. Then the total variation distance between η and ξ is defined as*

$$d_{TV}(\eta,\xi) = \sup_{\mathcal{N}\in\mathcal{B}\circ\mathcal{N}(S)} |\,\mathbb{P}\,(P\in\mathcal{N}) - \mathbb{P}\,(Q\in\mathcal{N})\,|,$$

where $\mathcal{L}(P)=\eta$ and $\mathcal{L}(Q)=\xi$.

We shall need another expression of the total variation metric, which is presented, for instance, in Theorem 5.2 of Lindvall [49] and obtained with coupling techniques.

Theorem 2.3.1 (γ-coupling) *Let η and ξ be distributions of elements of $\mathcal{P}(S)$, then*

$$d_{TV}(\eta,\xi) = \inf_{\substack{P:\mathcal{L}(P)=\eta\\ Q:\mathcal{L}(Q)=\xi}} \mathbb{P}\,(P\neq Q).$$

In our case, the problem with the total variation distance is that we compare a discrete point process to its asymptotic limit, with an intensity that is absolutely continuous with respect to Lebesgue measure, and realizations of two such random measures being equal with probability 0, their total variation distance would always be equal to 1. This is why a Wasserstein metric is more adequate in this framework; because it can compare random measures of different natures in a non-trivial way. Here is the Wasserstein metric that will be used in this study, as in Barbour *et al.* [7].

Definition 2.3.2 (Wasserstein distance) *Let η and ξ be distributions of elements of $\mathcal{P}(S)$. A Wasserstein distance between η and ξ has the following form.*

$$d_W(\eta,\xi) \;=\; \inf_{\substack{P:\mathcal{L}(P)=\eta \\ Q:\mathcal{L}(Q)=\xi}} \mathbb{E}\, d(P,Q)$$

where d is any distance between counting measures. The distance d we choose is the following, where $(N_1,N_2)\in\big(\mathcal{N}(S)\big)^2$,

$$d(N_1,N_2) \;=\; \begin{cases} 1 & \text{if } N_1(S)\neq N_2(S) \\ \Delta(N_1,N_2) & \text{if } N_1(S)=N_2(S)=N>0 \\ 0 & \text{if } N_1(S)=N_2(S)=0 \end{cases}$$

where the average distortion Δ *is defined as*

$$\Delta(N_1,N_2) \;=\; \frac{1}{N}\min_{\tau\in\Pi_N}\sum_{i=1}^{N}\min\{|k_{i,1}-k_{\tau(i),2}|,1\}.$$

Here Π_N is the set of permutations of $[\![1,N]\!]$, $k_{i,1}$ and $k_{j,2}$ are the indices of the vectors in the state space S by which the i^{th} point of N_1, respectively the j^{th} point of N_2, are carried. This Wasserstein distance can equivalently be defined as

$$d_W(\eta,\xi) \;=\; \sup_{f\in\mathcal{F}(S)}\left|\int f d\eta - \int f d\xi\right|$$

where $\mathcal{F}(S)$ is the set of functionals of counting measures, which are 1-Lipschitz with respect to the distance d specified above:

$$\mathcal{F}(S) \;=\; \big\{f:\mathcal{N}(S)\to\mathbb{R} : \forall\,(N_1,N_2)\in\big(\mathcal{N}(S)\big)^2, \big|f(N_1)-f(N_2)\big|\leq d(N_1,N_2)\big\}.$$

The average distortion between two counting measures having the same accumulated masses is the minimum average spatial distance between points (forming the mass carried by vectors of the state space) of each processes, *c.f.* Figure 2.1. The following corollary linking the two aforementioned metrics will be useful later, and its proof is straightforward given Theorem 2.3.1.

Corollary 2.3.1 *Let η and ξ be distributions of elements of $\mathcal{P}(S)$. Then, with the previous definition for d_W, $d_W(\eta,\xi)\leq d_{TV}(\eta,\xi)$.*

Proof.

$$
\begin{aligned}
& d_W(\eta,\xi) \\
&= \inf_{\substack{P:\mathcal{L}(P)=\eta \\ Q:\mathcal{L}(Q)=\xi}} \mathbb{E}\, d(P,Q) \\
&= \inf_{\substack{P:\mathcal{L}(P)=\eta \\ Q:\mathcal{L}(Q)=\xi}} \left\{ \mathbb{E}\left(d(P,Q)\middle|P=Q\right)\mathbb{P}\left(P=Q\right) + \mathbb{E}\left(d(P,Q)\middle|P\neq Q\right)\mathbb{P}\left(P\neq Q\right) \right\} \\
&= \inf_{\substack{P:\mathcal{L}(P)=\eta \\ Q:\mathcal{L}(Q)=\xi}} \left\{ \mathbb{E}\left(d(P,Q)\middle|P\neq Q\right)\mathbb{P}\left(P\neq Q\right) \right\} \\
&\leq \inf_{\substack{P:\mathcal{L}(P)=\eta \\ Q:\mathcal{L}(Q)=\xi}} \mathbb{P}\left(P\neq Q\right). \quad \square
\end{aligned}
$$

Remark. This is *not* true for all Wasserstein metrics. It is for our specific choice of the distance d.

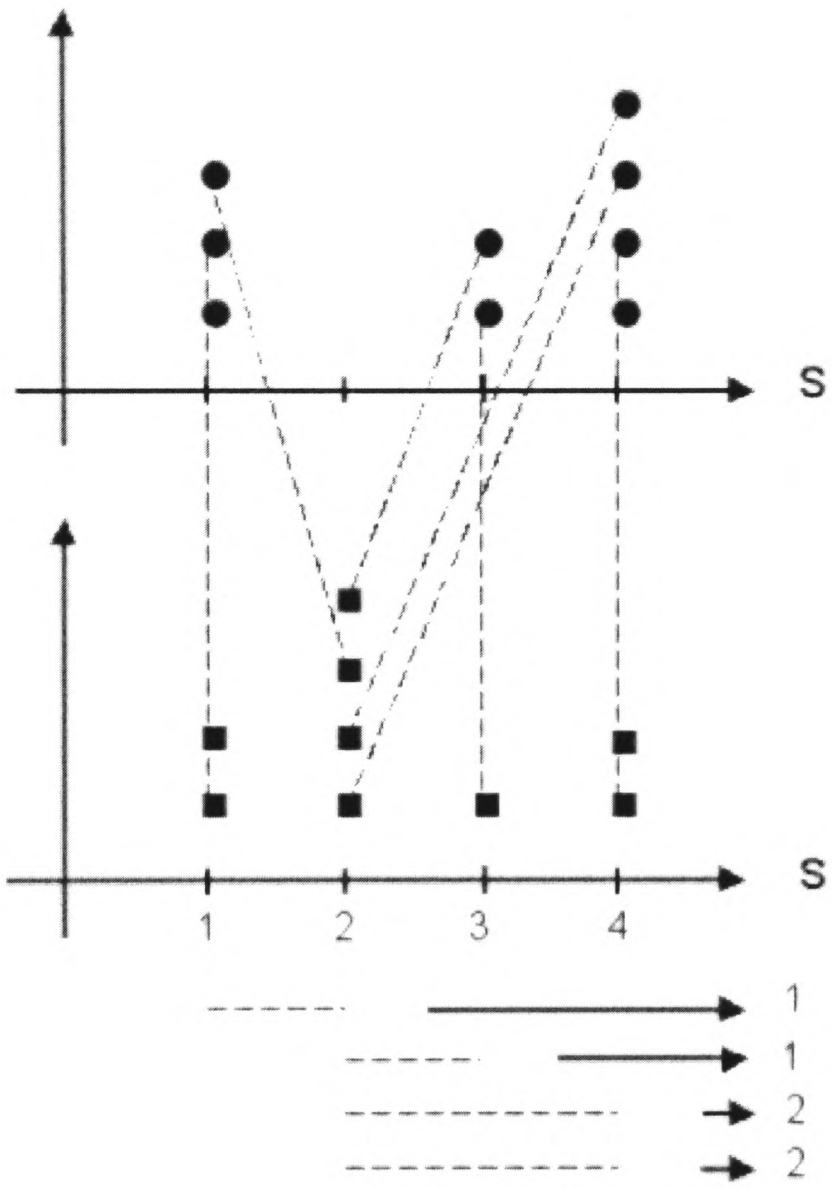

Figure 2.1: *Calculation of the average distortion between two counting measures having the same number of points over the whole state space. Here,* $S = [1, 4]$. *The bijection between the two sets of points drawn here minimizes the average distance between one point and its image. The corresponding average distance is the average distortion, here* $\Delta = (1 + 1 + 2 + 2)/9 = 6/9$.

2.4 Stability of stochastic processes

Now that we have presented the framework for the random measures we shall use, let us turn to the random variables we are going to employ as their scalar weights. We need here an incursion into the theory of Markov chains, in order to ready the tools to describe the time behaviours properly we shall be dealing with. Except where otherwise indicated, the following definitions are taken from Meyn and Tweedie [51], where mixing and stability properties of Markov chains are extensively presented. But before any Markovian consideration, it is fit to introduce the concept of stationarity.

Definition 2.4.1 (Stationarity) *A process* $(X_n)_{n\in\mathbb{N}^*}$ *on the state space* S *is* second-order stationary *if its mean is constant and if the covariance between* X_s *and* X_t *only depends on* $|s-t|$. *It is said to be* strictly stationary *if, given any time points* $t_1, \ldots, t_d$, *the joint distribution* $\mathcal{L}\big((X_{t_1+h}, \ldots, X_{t_d+h})\big)$ *is the same for any* $h\in\mathbb{N}^*$.

2.4.1 Markov chains

Now, a stochastic process $(X_t)_{t\in\mathbb{N}^*}$ on S is said to be a discrete time Markov chain when, for any measurable set $A \in \mathcal{B}(S)$, for any $d \in \mathbb{N}^*$, for any set of time points $t_1 < \cdots < t_d$, for any $(x_1, \ldots, x_{d-1}) \in S^{d-1}$, $\mathbb{P}(X_{t_d} \in A | X_{t_{d-1}} = x_{d-1}, \ldots, X_{t_1} = x_1) = \mathbb{P}(X_{t_d} \in A | X_{t_{d-1}} = x_{d-1})$. An invariant measure for the Markov chain is a σ-finite measure π on $\mathcal{B}(S)$ such that, for all $A \in \mathcal{B}(S)$, $\pi(A) = \int_S \pi(dx)\mathbb{P}(X_1 \in A | X_0 = x)$. Under certain regularity conditions, such an invariant measure is unique. In the case of a stationary Markov chain, it coincides with the stationary distribution. In view of the next chapters, let us examine the stability properties such chains can verify. First, the notion of irreducibility tells about how completely the chain covers the state space.

Definition 2.4.2 (φ-Irreducibility) *A Markov chain* $(X_t)_{t\in\mathbb{N}^*}$ *on* S *is* φ-irreducible *when there exists a measure* φ *on* $\mathcal{B}(S)$ *such that, for all* A *in* $\mathcal{B}(S)$ *verifying* $\varphi(A) > 0$,

for all $x \in S$, *there exists* $n \in \mathbb{N}^*$ *such that* $\mathbb{P}(X_n \in A | X_0 = x) > 0$.

We shall denote by ψ the "maximal" irreducibility measure of the Markov chain, *i.e.* the measure for which the chain is irreducible and such that all other irreducibility measures φ satisfy $\psi(A) = 0 \Longrightarrow \varphi(A) = 0$ for any $A \in \mathcal{B}(S)$. Such a unique measure ψ always exists for any φ-irreducible chain, and we shall use the classic notation ψ-*irreducible* for $(X_t)_{t\in\mathbb{N}^*}$ to mean that the irreducibility measure referred to is the "maximal" one. Now, because it is instrumental in the shaping of the concept of ergodicity, let us present the idea of a small set, singling out measurable subsets of the state space from which the chain can jump towards any direction in a uniformly finite number of steps.

Definition 2.4.3 (Small sets) *A set* $C \in \mathcal{B}(S)$ *is called a* small *set for* $(X_t)_{t\in\mathbb{N}^*}$ *if there exists an* $m \in \mathbb{N}^*$ *and a non-trivial measure* ν_m *on* $\mathcal{B}(S)$, *such that for all* $x \in C$, $B \in \mathcal{B}(S)$,

$$\mathbb{P}(X_m \in B | X_0 = x) \geq \nu_m(B).$$

C *is then said to be* ν_m-small.

A generalisation of this notion is obtained by looking at petite sets, where the fixed jump size above is replaced by a probability distribution of jump sizes.

Definition 2.4.4 (Petite sets) *A set* $C \in \mathcal{B}(S)$ *is called a* petite *set for* $(X_t)_{t\in\mathbb{N}^*}$ *if there exists a probability distribution* a *on* $\mathbb{Z}^+$ *and a non-trivial measure* ν_a *on* $\mathcal{B}(S)$, *such that for all* $x \in C$, $B \in \mathcal{B}(S)$,

$$\sum_{n=0}^{+\infty} \mathbb{P}(X_n \in B | X_0 = x) a(n) \geq \nu_a(B).$$

C *is then said to be* ν_a-petite.

Clearly, a ν_m-small set is ν_{δ_m}-petite. We are now going to define the notion of period for a Markov chain. First, let us take the simpler case of a countable state space. Let $x \in S$. Let $d(x)$ be the greater common divisor of the set $\{n \geq 1 :$

$\mathbb{P}(X_n = x|X_0 = x) > 0\}$. Then $d(x)$ is called the *period* of x. The chain $(X_t)_{t\in\mathbb{N}^*}$ is called *aperiodic* if $\forall x \in S, d(x) = 1$. Now let us turn to the case of a general state space S. Let C be a ν_M-small set such that $\nu_M(C) > 0$. Hence there is a positive probability that the chain return to C after M lags. Let $E_C = \{n \geq 1 : \exists \delta_n > 0 :$ C is $(\delta_n \nu_M)$-small$\}$. Let $d(C)$ be the greatest common divisor of E_C. It is shown in the proof of Theorem 5.4.4 in Meyn and Tweedie [51] that $d(C)$ is independent of the small set C chosen. When $d = 1$, the chain $(X_t)_{t\in\mathbb{N}^*}$ is called *aperiodic*. Let us now introduce a fundamental concept of stochastic stability, describing the convergence of a Markov chain to its invariant distribution.

Definition 2.4.5 (Uniform ergodicity) *A Markov chain $(X_t)_{t\in\mathbb{N}^*}$ on S with invariant distribution π is* uniformly ergodic *when*

$$\sup_{x\in S} d_{TV}\Big(\mathbb{P}(X_n \in \bullet|X_0 = x), \pi(\bullet)\Big) \underset{n\to+\infty}{\longrightarrow} 0.$$

Theorem 16.2.2 (*ibid.*) states the equivalence between the petiteness of the state space S and the uniform ergodicity of $(X_t)_{t\in\mathbb{N}^*}$, in the case when $(X_t)_{t\in\mathbb{N}^*}$ is ψ-irreducible and aperiodic. This criterion is easy to manipulate for our perpetuity. Now, for $A \in \mathcal{B}(S)$, we shall denote the *occupation time* of A by $\eta_A = \sum_{n=1}^{+\infty} \mathbb{I}_{\{X_n\in A\}}$, *i.e.* the number of visits by the chain to A after time 0. A set $A \in \mathcal{B}(S)$ is called *Harris recurrent* if for all $x \in A$, $\mathbb{P}(\eta_A = +\infty|X_0 = x) = 1$. A Markov chain is called *Harris recurrent* if it is ψ-irreducible and every set $A \in \mathcal{B}(S)$ such that $\psi(A) > 0$ is Harris recurrent. Theorem 3.3 in Meyn and Tweedie [52] shows the equivalence between the Harris recurrence of a Markov chain and the existence of a petite set C such that for all $x \in C$, $\mathbb{P}(\tau_C < +\infty|X_0 = x) = 1$, where the *first return time* of C is denoted by $\tau_C = \min\{n \geq 1 : X_n \in C\}$. Davydov [26] showed that Harris recurrent Markov chains are ergodic and β-mixing (*c.f.* next section).

2.4.2 Absolute regularity

The notion of ergodicity is very powerful, and creates interesting time-dependence features in a stochastic process. A way of looking into this is to assess its mixing behaviour. Mixing can be analysed through various coefficients. In this work, we shall essentially be dealing with β-mixing, also called absolute regularity, because it yields key properties that will be used in the following chapters. We refer to Doukhan [27] for the related definitions.

Definition 2.4.6 (β-mixing) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\mathcal{U}$ and $\mathcal{V}$ be two sub σ-fields of $\mathcal{A}$. The* β-mixing *or* absolute regularity *coefficient of $\mathcal{U}$ and $\mathcal{V}$ is defined as*

$$\beta(\mathcal{U}, \mathcal{V}) = \mathbb{E} \sup_{U \in \mathcal{U}} \big| \mathbb{P}(U|\mathcal{V}) - \mathbb{P}(U) \big|.$$

It can equivalently be rewritten as

$$\beta(\mathcal{U}, \mathcal{V}) = \sup \left\{ \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \big| \mathbb{P}(U_i)\mathbb{P}(V_j) - \mathbb{P}(U_i \cap V_j) \big| \right\},$$

where the supremum is taken over all the partitions $(U_i)_{i \in I}$, $(V_j)_{j \in J}$ of Ω with $U_i \in \mathcal{U}$ and $V_j \in \mathcal{V}$. Consequently, this latter expression shows that β is symmetric. Now, if $(X_t)_{t \in \mathbb{N}^*}$ is a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$, let us denote its *absolute regularity coefficient* by

$$\beta(k) = \sup_{t \in \mathbb{N}^*} \beta\big(\sigma(X_s, s \leq t), \sigma(X_s, s \geq t+k)\big).$$

The process $(X_t)_{t \in \mathbb{N}^*}$ is said to be *β-mixing* or *absolutely regular* if $\beta(k)$ tends to 0 as k tends to infinity. For a finite sequence $(X_t)_{1 \leq t \leq n}$, we denote

$$\beta_X(k) = \max_{1 \leq t \leq n} \beta\big(\sigma(X_s, s \leq t), \sigma(X_s, t+k+1 \leq s \leq n)\big).$$

Let us now relate this notion to the general Markovian properties exposed in the previous section, through a certain type of ergodicity. A Markov chain $(X_t)_{t \in \mathbb{N}^*}$ is

said to be *geometrically ergodic* if it has an invariant probability measure π and there are two constants $\rho \in]0,1[$ and $R > 0$ such that, for all $x \in S$, $d_{TV}\Big(\mathbb{P}(X_n \in \bullet | X_0 = x), \pi(\bullet)\Big) \leq R\rho^n$. Now, Davydov [26] provides a useful expression of the absolute regularity coefficient. Indeed, if $(X_t)_{t\in\mathbb{N}^*}$ is a homogeneous Markov chain on $(S, \mathcal{B}(S))$, then

$$\beta(l) = \sup_{t\in\mathbb{N}^*} \int_S \mathbb{P}(X_t \in dx) d_{TV}\Big(\mathbb{P}(X_l \in \bullet | X_0 = x), \mathbb{P}(X_{t+l} \in \bullet)\Big).$$

Now, if the process of interest also happens to be stationary, with invariant distribution denoted by π, we get

$$\beta(l) = \int_S \pi(dx) d_{TV}\Big(\mathbb{P}(X_l \in \bullet | X_0 = x), \pi(\bullet)\Big).$$

Hence if our process is geometrically ergodic, then we have the existence of two constants $\rho \in]0,1[$ and $R > 0$ such that $\beta(l) \leq R\rho^l$. Now there is another fact about absolute regularity coefficients that will be precious at a later stage. Let us present it in the following lemma.

Lemma 2.4.1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $\mathcal{U}$ and $\mathcal{V}$ be two sub σ-fields of $\mathcal{A}$. Let $\mathcal{U}^*$ and $\mathcal{V}^*$ in $\mathcal{A}$ such that $\mathcal{U}^* \subset \mathcal{U}$ and $\mathcal{V}^* \subset \mathcal{V}$. Then $\beta(\mathcal{U}^*, \mathcal{V}^*) \leq \beta(\mathcal{U}, \mathcal{V})$.*

Proof.

$$\begin{aligned} \beta(\mathcal{U}^*, \mathcal{V}^*) &= \mathbb{E} \sup_{U\in\mathcal{U}^*} \big| \mathbb{P}(U|\mathcal{V}^*) - \mathbb{P}(U) \big| \\ &\leq \mathbb{E} \sup_{U\in\mathcal{U}} \big| \mathbb{P}(U|\mathcal{V}^*) - \mathbb{P}(U) \big|, \end{aligned}$$

because $\mathcal{U}^* \subset \mathcal{U}$. Hence $\beta(\mathcal{U}^*, \mathcal{V}^*) \leq \beta(\mathcal{U}, \mathcal{V}^*)$. Now, by symmetry, $\beta(\mathcal{U}^*, \mathcal{V}^*) \leq \beta(\mathcal{V}^*, \mathcal{U})$. In other words,

$$\begin{aligned} \beta(\mathcal{U}, \mathcal{V}^*) &= \mathbb{E} \sup_{V\in\mathcal{V}^*} \big| \mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V) \big| \\ &\leq \mathbb{E} \sup_{V\in\mathcal{V}} \big| \mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V) \big|, \end{aligned}$$

because $\mathcal{V}^* \subset \mathcal{V}$. Hence the result. □

Remark. As a result, if the process is geometrically ergodic, then the β-mixing

coefficient $\beta_X(l)$ associated with a finite sequence from $(X_t)_{t\in\mathbb{N}^*}$ is also bounded by $R\rho^l$, because $\sigma(X_s, t+k+1 \leq s \leq n) \subset \sigma(X_s, s \geq t+k+1)$. In addition, when we focus on the exceedance process $\left(\mathbb{I}_{\{X_t>u\}}\right)_{1\leq t\leq n}$, denoting by $\beta_u(l)$ its mixing coefficient, given that we have the following equality

$$\beta_u(l) \;=\; \beta\Big(\sigma\big(\{X_k > u\}, 1 \leq k \leq t\big), \sigma\big(\{X_k > u\}, t+m+1 \leq k \leq n\big)\Big),$$

and noting that $\sigma\big(\{X_k > u\}, s \leq k \leq t\big) \subset \sigma(X_k, s \leq k \leq t)$, we also have $\beta_u(l) \leq R\rho^l$.

Chapter 3

Extreme value theory

This thesis is about the exceedance behaviour of stochastic recursions, using tools recently developed in extreme value theory, so it is now time to introduce a few related notions, and set up the extreme value framework. One of the important results concerns the way extremes react to time correlation in the original time series. More precisely, extreme value theory provides the parameters of the asymptotic compound Poisson approximation for the exceedance point process. Indeed, there are several ways of assessing extremal behaviours of stationary sequences. I have chosen to look at these extremal behaviours from the viewpoint of exceedances $\mathbb{I}_{\{X_t \geq u\}}$ above a threshold u, which asymptotically form a compound Poisson process. Other approaches are available in the extreme value theory literature, such as sequence maxima $M_t = \max\{X_s, 1 \leq s \leq t\}$, with a generalised extreme value limit distribution, or threshold excesses $Y_t = \max\{X_t - u, 0\}$, with a a generalised Pareto limit distribution. Further information on these methods can be found in Embrechts *et al.* [30]. Let us first introduce the notion of extremal index.

3.1 Extremal index

First, we need to state a fundamental result on the behaviour of extremes, in the simple, i.i.d. case, *c.f.* Theorem 3.2.3 in Embrechts *et al.* [30].

Theorem 3.1.1 (Fisher-Tippett theorem) *Let $(X_n)_{n\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables. If there are norming constants $(c_n, d_n)_{n\in\mathbb{N}^*}$ and some non-degenerate distribution H such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow[n\to+\infty]{d} H \tag{3.1}$$

then H is a generalised extreme value distribution, of the form

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}} & \text{for } 1+\xi x > 0, \text{ if } \xi \neq 0, \\ e^{-e^{-x}} & \text{for } x > 0, \text{ if } \xi = 0. \end{cases}$$

Here, $\xi = \frac{1}{\alpha} > 0$ corresponds to the Fréchet distribution ϕ_α, $\xi = 0$ corresponds to the Gumbel distribution Λ, and $\xi = -\frac{1}{\alpha} < 0$ corresponds to the Weibull distribution Ψ_α.

This major result gives the complete shape of the sequence maxima asymptotic distribution. Now, what if we allow for time correlation in the original $(X_n)_{n\in\mathbb{N}^*}$ series? Here comes in the notion of *extremal index*, *c.f.* Leadbetter *et al.* [47].

Definition 3.1.1 (Extremal index) *A second-order stationary sequence $(X_n)_{n\in\mathbb{N}^*}$ of random variables is said to have extremal index $\theta \in [0,1]$ if*

$$\forall \tau \in \mathbb{R}^+, \exists \left(u_n(\tau)\right)_{n\in\mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*} : \begin{cases} n\overline{F}\left(u_n(\tau)\right) \xrightarrow[n\to+\infty]{} \tau, \\ \mathbb{P}\left(M_n \leq u_n(\tau)\right) \xrightarrow[n\to+\infty]{} e^{-\theta\tau}. \end{cases}$$

Many stationary processes have an extremal index, for instance linear processes driven by a noise satisfying certain regularity conditions, but not all of them have, *c.f.* Examples 8.1.1 and 8.1.4 respectively in Embrechts *et al.* [30].

Theorem 3.1.2 *Let $(X_n)_{n\in\mathbb{N}^*}$ be a strictly stationary time series with extremal index*

θ, let $(\widetilde{X}_n)_{n\in\mathbb{N}^*}$ be i.i.d., with $\widetilde{X}_1 \stackrel{d}{=} X_1$. If (3.1) holds for some H, then

$$\frac{M_n - d_n}{c_n} \xrightarrow[n\to+\infty]{d} H^\theta \iff \frac{\widetilde{M}_n - d_n}{c_n} \xrightarrow[n\to+\infty]{d} H$$

The extremal index is the consequence on the extremes of a deviation from independence in the original time series. Yet, there is another way to look at it, through the point process approach.

3.2 Threshold exceedances

Yet another essential result for us concerns the asymptotic behaviour of empirical point processes of exceedances. It can be found in Embrechts *et al.* [30], and is an application of Kallenberg's theorem for weak convergence to a simple point process on an interval.

Theorem 3.2.1 (Convergence of exceedance point process) *Let $(X_n)_{n\in\mathbb{N}^*}$ be an i.i.d. sequence of random variables with margins F. Let $(u_n)_{n\in\mathbb{N}^*}$ be a sequence of threshold values in $\mathbb{R}$, such that*

$$\exists \tau \in \mathbb{R} : \quad n\overline{F}(u_n) \xrightarrow[n\to+\infty]{} \tau$$

Then, if the state space is $]0,1]$,

$$N_n(\bullet) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i > u_n\}} \delta_{i/n}(\bullet) \xrightarrow[n\to+\infty]{d} \text{Poisson}\big(\tau \boldsymbol{\mu}(\bullet)\big)$$

where $\boldsymbol{\mu}$ is the Lebesgue measure on $]0,1]$.

If we allow for time correlation in the original empirical point process of exceedances, it will asymptotically form a *compound* Poisson random measure, as we shall see later. The latter models are a mixture of a homogeneous Poisson process of clusters compounded with i.i.d. cluster sizes. A spectacular result is that *the limiting mean cluster size is the inverse of the extremal index* of the original sequence, *c.f.*

Embrechts *et al.* [30]. This is a feature we shall also come across in the course of the empirical modelling, in Section 5. Looking at the size of exceedance clusters is therefore another way to approach the extremal index.

3.3 Multivariate case

The asymptotic theory for exceedance point processes in a multivariate setting is developed in Davis and Hsing [24]. They show that under appropriate assumptions of mixing and regular variation, a certain point process of exceedances on $\mathbb{R}^d$ converges to a compound Poisson measure. This is summarised in Davis and Mikosch [25]. Let us be more specific. Let $(\boldsymbol{X}_t)$ be a strictly stationary stochastic process on $\mathbb{R}^d$, with the max-norm $\|\cdot\|_\infty$, *i.e.* if $\boldsymbol{x} = (x_1, \ldots, x_d)$, then $\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq d} |x_i|$. First, we shall assume that all the covariates $(X_1, \ldots, X_d)$ are *regularly varying* with tail index α, which means that there exists $\alpha > 0$ such that

$$\forall\, i \in [\![1, d]\!], \frac{\mathbb{P}\big(|X_i| > tx\big)}{\mathbb{P}\big(|X_i| > x\big)} \xrightarrow[x \to +\infty]{v} t^{-\alpha},$$

where $\xrightarrow{v}$ denotes the vague convergence on $\mathbb{S}^{d-1}$. Recall that a sequence of random measures μ_n converges vaguely to a random measure μ if $\int_S g(x)\mu_n(dx)$ tends to $\int_S g(x)\mu(dx)$ as $n \longrightarrow +\infty$ for any continuous, non-negative function g with compact support. In addition, we shall assume that the stationary distribution of $\boldsymbol{X}$ is *regularly varying* with index α and spectral measure $\mathbb{P}(\boldsymbol{\theta} \in \bullet)$. This means that, in addition to the α mentioned above, there exists a random vector $\boldsymbol{\theta}$ almost surely in the $\|\cdot\|_\infty$-unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ such that, for all $t > 0$,

$$\frac{\mathbb{P}\big(\|\boldsymbol{X}\|_\infty > tx, \boldsymbol{X}/\|\boldsymbol{X}\|_\infty \in \bullet\big)}{\mathbb{P}\big(\|\boldsymbol{X}\|_\infty > x\big)} \xrightarrow[x \to +\infty]{v} t^{-\alpha}\mathbb{P}(\boldsymbol{\theta} \in \bullet),$$

In such a case, the factorisation above can be achieved for any norm on $\mathbb{R}^d$, although the form of $\mathbb{P}(\boldsymbol{\theta} \in \bullet)$ will vary according to the norm chosen. Further information on regular variation can be found in Bingham *et al.* [15]. Let us now turn to the second

assumption that is necessary to the convergence of our point process. Let (a_n) be a sequence of positive numbers such that

$$n\mathbb{P}\big(\|\boldsymbol{X}\|_\infty > a_n\big) \underset{n\longrightarrow+\infty}{\longrightarrow} 1.$$

These are anologous to our thresholds u_n in Theorem 3.2.1, and this is the type of exceedance we are looking at in this particular multivariate setting. For a sequence of given length n, we look at how likely it is that $\|\boldsymbol{X}\|_\infty$ be bigger than a_n. In other words, we are considering exceedances outside a hypercube centered on the origin and of volume a_n^d. Still following Davis and Mikosch [25], we shall say that Condition $\mathcal{A}(a_n)$ holds if there exists a sequence of positive integers (r_n), such that $r_n \to +\infty$ and $\lfloor n/r_n \rfloor \to +\infty$ as $n \to +\infty$, and for all bounded non-negative step function f on $\overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$ with bounded support,

$$\mathbb{E}\exp\Big(-\sum_{t=1}^{n} f(\boldsymbol{X}_t/a_n)\Big) - \Big(\mathbb{E}\exp\Big(-\sum_{t=1}^{r_n} f(\boldsymbol{X}_t/a_n)\Big)\Big)^{\lfloor n/r_n\rfloor} \underset{n\to+\infty}{\longrightarrow} 0.$$

This is a very weak condition, implied by many classic mixing conditions, which was introduced in Davis and Hsing [24]. It is also independent of the sequence (a_n) chosen. Now, if the stricly stationary process $(\boldsymbol{X}_t)$ is regularly varying and verifies the condition $\mathcal{A}(a_n)$, then the point process

$$N_n(\bullet) \;=\; \sum_{t=1}^{n} \mathbb{I}_{\{\boldsymbol{X}_t/a_n \in \bullet\}}$$

converges in distribution to a limiting compound Poisson measure. The Poisson intensity is given by $\lambda(dy) = \theta\alpha y^{-\alpha-1}\mathbb{I}_{\{y\geq 0\}}dy$, where α is the coefficient of regular variation, and θ is the extremal index of the sequence $(\|\boldsymbol{X}_t\|_\infty)$. The detailed expressions of the compound Poisson parameters can be found in Theorem 2.7 of Davis and Hsing [24] in the form of asymptotic limits. The present work is partly motivated by a desire to find different ways of describing threshold exceedances, allowing easier calculations.

Now, for the joint distribution of covariates, the theory of regular variations provides an interesting framework to describe the vector in an extreme situation.

First, the index α gives an indication of the approximate length of the vector, and then its spectral measure tells us about its angle, *i.e.* about the dependence bonds between its covariates when in an extreme situation. There is *asymptotic dependence* between the extremes of the covariates whenever the limit distribution of $\boldsymbol{\theta}$ has some mass outside the origin axes. For example, $\boldsymbol{\theta} = \pi/4$ corresponds to a situation when the extremes of the covariates are asymptotically exactly dependent if all the mass is concentrated on that value, and when $\boldsymbol{\theta} = \pi/2$ or $\boldsymbol{\theta} = 0$, extremal events are asymptotically independent across covariates. A consistent estimator for the spectral measure can be found in Stărică [64] for the case of joint extremal events, when all the covariates take large values at the same time. Now, the joint distribution can also be examined from the viewpoint of componentwise maxima. In this approach, the marginal lengths of the covariates are described by the Fisher-Tippett theorem (*c.f.* Theorem 3.1.1), and the dependence structure by copulas, as in Embrechts *et al.* [31] and Capéraà *et al.* [20]. Now, these two approaches only lead to an efficient estimation of the joint distribution when all components are jointly large. Overcoming this restriction, Heffernan and Tawn [40] use a conditional approach, noting that all covariates need not be unusually large for the extreme value theory framework to apply.

Chapter 4

First illustration of the whole approximation

An analysis of threshold exceedances in the log-return process of Sterling *vs* Dollar exchange rates from 2 November 1992 to 15 June 2001 can be found in Benjamin [13]. This study, by the issues it raised, triggered the idea of this thesis. Full details of the analysis are in Benjamin [13], and in this chapter I shall only give a brief outline of it, in order to provide the reader with an intuitive idea of the approximation method developed in later chapters. The first part is a classical Box-Jenkins cleaning of the data downloaded from the Federal Reserve website (http://www.federalreserve.gov/Releases/h10/Hist/), *c.f.* Figure 4.1, in order to obtain a stationary series. Then, ARIMA and GARCH models are fitted to it, and a GARCH(1,1) model is selected. Then the focus is put on threshold exceedances. To have a rough, inaccurate idea on the accuracy of the compound Poisson approximation, the parameters of the asymptotic model are calculated, according to Mikosch

and Stărică [53], and Goldie [36]. The count of exceedances is then compared to its limiting expected value, and the bound given in Barbour *et al.* [7] is calculated and compared to these two quantities. As will be explained later on, these comparisons are not mathematically rigorous, for quantities are compared that should not be. They have the sole purpose to give an intuitive idea of the situation.

4.1 Sterling/Dollar exchange rates

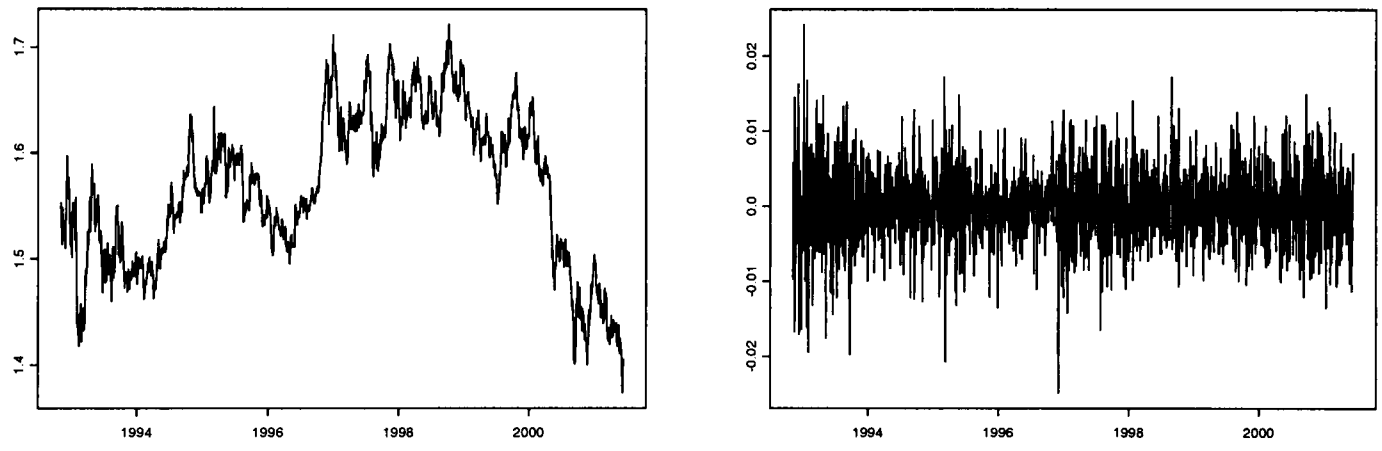

Figure 4.1: *On the left, Sterling/Dollar exchange rates from 02/11/1992 to 15/06/2001. On the right, the series of the detrended and deseasonalised log-returns, without any significant non-stationary components.*

The original time series $(Y_t)_{1\leq t\leq 2248}$ is "cleaned" to obtain a stationary one. For that purpose, the classical Box-Jenkins approach is applied to the log-returns $X_t = \log Y_t - \log Y_{t-1}$ of the data. The remaining trend is removed with a generalised least square analysis, and the seasonal components are removed according to Cleveland *et al.* [21]. Now, there is the issue of the noise that might have accumulated during weekends. Indeed, studying the process as a regularly spaced time series with weeks of five days might be questionable, as noises do not stop their evolution during the weekend. We shall assume that this can be neglected once all significant seasonal

components have been removed, *c.f.* Cleveland *et al.* [21] for a broader discussion on calendar adjustment. The issue of data missing because of bank holidays has been addressed. There are so few of them that we can safely impute them, taking the median of the four closest neighbours, without artificially reducing the overall variance. Models among the ARIMA and GARCH families are then fitted to the resulting series, and the best of them selected by Akaike's Information Criterion and a Bayesian Information Criterion. If there had not been any indication of heteroscedasticity, an MA(1) model would have been chosen, but the analysis of residuals showed that such a fit was poor. Finally, a GARCH(1,1) model, described in the next section, is chosen.

4.2 GARCH models

Autoregressive conditionally heteroscedastic (ARCH) models have been introduced in the early 1980's, when the increasing volatility of markets made it clear there was much to explore in the stochastic behaviour of variances, *c.f.* Engle [32]. A stochastic recursion was set up to explain the conditional variance at time t by the value of the return at time $t-1$. A generalization (GARCH) was then introduced when the conditional variance at time $t-1$ was added as an explanator, *c.f.* Bollerslev [16].

Definition 4.2.1 (GARCH(1,1)) *A sequence $(X_t)_{t\in\mathbb{N}^*}$ forms a generalized autoregressive conditionally heteroscedastic (GARCH) process of order (1,1) if and only if there exist a sequence $(Z_t)_{t\in\mathbb{N}^*}$ of i.i.d. symmetric random variables, called innovations, and a sequence $(\sigma_t)_{t\in\mathbb{N}^*}$ of random variables, called volatilities, such that*

$$\forall t\in\mathbb{N}^*, \quad X_t = \sigma_t Z_t, \tag{4.1}$$

where σ_t verifies

$$\forall t\in\mathbb{N}^*, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{4.2}$$

where α_0, α_1 and β_1 are constants. The conditions for the existence of a stationary solution to (4.2) are that $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and that in addition $\alpha_1 + \beta_1 < 1$.

In our particular case, a $t_{6.84}$ distribution is selected for the innovations. The number of degrees of freedom of 6.84 is estimated from the data. The following approximate 95% maximum likelihood confidence intervals are derived for the GARCH(1,1) coefficients, jointly estimated, even if no covariance matrix is provided here. The intervals do not correspond to the marginal distribution of each parameter, and therefore should not be read independently of each other.

$$\alpha_0 = 1.331 \times 10^{-7} \in [9.857 \times 10^{-9}, 2.563 \times 10^{-7}]$$
$$\alpha_1 = 2.482 \times 10^{-2} \in [1.595 \times 10^{-2}, 3.369 \times 10^{-2}]$$
$$\beta_1 = 9.573 \times 10^{-1} \in [9.411 \times 10^{-1}, 9.734 \times 10^{-1}]$$

4.3 Marginal tail

Theorem 3.2.1 means that, in the case of an i.i.d. sequence, we need to know the marginal tail in order to find out the intensity of the asymptotic Poisson process of threshold exceedances. Here, our GARCH sequence is clearly not i.i.d., and the limit of the empirical point process of exceedances is not a Poisson, but a compound Poisson measure. Yet, in Mikosch and Stărică [53], a similar tail argument is developed to find the intensity and compounding distribution. The results on the tail of a perpetuity used here are fully presented in Theorem 7.2.1 of Chapter 7.

Equation (4.2) is a stochastic recurrence equation in σ^2, therefore we are in a position to calculate the shape of its marginal tail, namely, c and κ of Theorem 7.2.1. The appropriate κ is the non-zero solution of

$$\mathbb{E}\left(\alpha_1 Z^2 + \beta_1\right)^\kappa = 1,$$

where Z is a $t_{6.84}$ random variable. κ is computed numerically in Benjamin [13]

($\kappa \approx 3.51$). In *ibid.* [13], an application of results from Goldie [36] to this particular case gives the following value for c

$$c = \frac{\mathbb{E}\left[(\alpha_0 + (\alpha_1 Z^2 + \beta_1)\sigma^2)^\kappa - ((\alpha_1 Z^2 + \beta_1)\sigma^2)^\kappa\right]}{\kappa\, \mathbb{E}\left[(\alpha_1 Z^2 + \beta_1)^\kappa \ln(\alpha_1 Z^2 + \beta_1)\right]}.$$

Yet, we are still not dealing with the shape of the log-return tail itself. This is achieved by first noting that we have

$$\forall t \in \mathbb{N}^*, \quad \mathbb{P}(\sigma_t^2 > x) \underset{x \to +\infty}{\sim} c x^{-\kappa},$$

which implies

$$\forall t \in \mathbb{N}^*, \quad \mathbb{P}(\sigma_t > x) \underset{x \to +\infty}{\sim} \frac{c}{2} x^{-2\kappa}.$$

Indeed this follows from Theorem 7.2.1 applied to $Y_t = \sigma_t^2$, $B_t = \alpha_1 Z_t^2 + \beta_1$, and $A_t = \alpha_0$. In addition, from Mikosch and Stărică [53], where threshold exceedances of a GARCH(1,1) process are comprehensively investigated, we have

$$\mathbb{P}(|X| > x) = \mathbb{E}|Z|^{2\kappa}\, \mathbb{P}(\sigma > x),$$

where X, σ and Z respectively have the stationary distributions associated with $(X_t)_{t \in \mathbb{N}^*}$, $(\sigma_t)_{t \in \mathbb{N}^*}$ and $(Z_t)_{t \in \mathbb{N}^*}$ from (4.1). We now know the tail shape of $|X|$. This enables us to derive the asymptotic compound Poisson distribution of the threshold exceedances in $(X_t)_{t \in \mathbb{N}^*}$.

4.4 Compound Poisson model

The following convergence theorem, analogous to Theorem 3.2.1, is from Mikosch and Stărică [53], as a variation on results from Davis and Hsing [24].

Theorem 4.4.1 *Let $(X_n)_{n \in \mathbb{N}^*}$ be a sequence of random variables. If $(a_n)_{t \in \mathbb{N}^*}$ verifies*

$$n\, \mathbb{P}(|X| > a_n) \underset{n \to +\infty}{\longrightarrow} 1$$

then, on the space $]0,1]$,

$$N_n(\bullet) = \sum_{k=1}^{n} \mathbb{I}_{\{X_k > x a_n\}} \delta_{k/n}(\bullet) \xrightarrow[n\to+\infty]{d} \mathrm{CP}\big(\theta_{|X|} x^{-2\kappa} \boldsymbol{\mu}(\bullet), \nu\big),$$

where $\theta_{|X|}$ *is the extremal index of the sequence* $(|X_t|)_{t\in\mathbb{N}^*}$, $\boldsymbol{\mu}(\bullet)$ *is the Lebesgue measure on* $]0,1]$, *and* ν *is a probability distribution on* $\mathbb{R}_+$.

This means that the point process of exceedances above xa_n within $\{X_1, \ldots, X_n\}$ tends, as n tends to infinity, to a compound Poisson measure, composed of

- a homogeneous Poisson process of intensity $\theta_{|X|} x^{-2\kappa}$ corresponding to the occurrence of clusters
- the compounding distribution ν of the i.i.d. cluster sizes, with mean $1/\theta_{|X|}$.

4.5 Bound on the approximation

Barbour, Novak and Xia [7] have found a Wasserstein bound on a compound Poisson approximation for the distribution of threshold exceedances, in the general case of strictly stationary sequences. The aim of this first illustration is to compare it to data, and have an idea about its sharpness. However, the approximating model used is empirical, and not the asymptotic one we mentioned before. Let us first present this empirical model.

4.5.1 Method of Bernstein blocks

The parameters of this empirical approximation are dependent on an arbitrary chopping of the sequence into blocks of size r. This is the Bernstein blocks method, popular in the Russian school, *c.f.* Ibragimov [42]. Our sequence is thence chopped into $\lfloor \frac{n}{r} \rfloor$ blocks of equal size r, plus a last one of size $n - r\lfloor \frac{n}{r} \rfloor$. The idea is to look

at the sequence of numbers of exceedances per block. Under suitable mixing considerations, and with a judicious choice of r, this sequence should be close to i.i.d., and when it actually is i.i.d., then a result by Xia [68], using Stein's method, enables us to make the link with the approximating compound Poisson process.

Let us be more precise. We have our finite, strictly stationary sequence $(X_i)_{1\leq i\leq n}$ of real-valued random variables, and our real threshold u. An exceedance is an event $\{X_i > u\}$. For $i \in \{1, \ldots, \lceil \frac{n}{r} \rceil\}$, the i^{th} Bernstein block of size r is the set of indices $B_r(i) = \{(i-1)r + 1, \ldots, (ir) \wedge n\}$. We denote by $T_{r,i}$ the number of exceedances in $B_r(i)$, *i.e.*

$$T_{r,i} = \sum_{k \in B_r(i)} \mathbb{I}_{\{X_k > u\}} \tag{4.3}$$

An important point to bear in mind is that we are about to compare our exceedance point process, composed of points in the state space that are typically arbitrarily spaced by units of 1 on the real line, to a random measure with mean intensity $p = \mathbb{P}(X_1 > u)$, with the Euclidean distance contained in the definition of our Wasserstein metric. If we do not do anything about this, we do not have scale invariance. This is why we shall stretch the state space to make the intensity 1, so that the Euclidean comparison between the processes is meaningful. Hence, the quantity we are approximating is the empirical point process of exceedances on the state space $[0, np]$:

$$N_n(\bullet) = \sum_{i=1}^{n} \mathbb{I}_{\{X_i > u\}} \delta_{ip}(\bullet)$$

and the intermediate variable we are using is the following empirical point process, on the same state space $[0, np]$:

$$\sum_{i=1}^{\lceil \frac{n}{r} \rceil} T_{r,i} \delta_{irp}(\bullet)$$

where the points of a block $B_r(i)$ have been gathered in its endpoint $\delta_{ir}(\bullet)$. Denoting by $\theta_r = \frac{1}{rp} \mathbb{P}(T_{r,1} \geq 1)$, $\nu_r = \mathcal{L}(T_{r,1} | T_{r,1} \geq 1)$ and $\boldsymbol{\mu}$ the Lebesgue measure, the

approximating compound Poisson process we shall use is

$$\boxed{\mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)}.$$

4.5.2 Link with the asymptotic model

Under the following convergence conditions, with sequences of thresholds $(u_n)_{n\in\mathbb{N}^*}$ and of block sizes $(r_n)_{n\in\mathbb{N}^*}$,

- $n\,\mathbb{P}\,(X>u_n)$ is bounded away from 0 and $+\infty$
- $r_n \to \infty$ and $r_n/n \to 0$,

there exists a limit

$$\theta = \lim_{n\to+\infty} \theta_{r_n}$$

which is the extremal index of the sequence $(X_n)_{n\in\mathbb{N}^*}$, *c.f.* O'Brien [56]. Recall from Section 3 that the extremal index is the inverse of the mean cluster size. Following is a suggested proof to show that our empirical parameters θ_r and ν_r already carry this feature. As ν_r is the empirical compounding distribution, the mean cluster size is $\mathbb{E}\,\nu_r$, and

$$\begin{aligned}
\mathbb{E}\,\nu_r &= \mathbb{E}\,(T_{r,1}|T_{r,1}\geq 1) \\
&= \sum_{i=1}^{r} i\,\mathbb{P}\,(T_{r,1}=i|T_{r,1}\geq 1) \;=\; \frac{1}{\mathbb{P}\,(T_{r,1}\geq 1)} \sum_{i=1}^{r} i\,\mathbb{P}\,(T_{r,1}=i) \\
&= \frac{\mathbb{E}\,T_{r,1}}{\mathbb{P}\,(T_{r,1}\geq 1)} \;=\; \frac{rp}{\mathbb{P}\,(T_{r,1}\geq 1)} \;=\; \frac{1}{\theta_r},
\end{aligned}$$

so our empirical parameters θ_r and ν_r already behave like the asymptotic ones.

4.5.3 Heuristic assessment

The bound given in Theorem 2.1 of Barbour *et al.* [7] has the following form. Let $n\in\mathbb{N}^*$. If $(X_i)_{1\leq i\leq n}$ is a strictly stationary sequence of random variables, then

$$d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r\boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \leq \epsilon(n,r,m) \tag{4.4}$$

with the metric d_W of Section 2.3, and we are looking for convenient ways of assessing its sharpness. The bound ϵ is a function of the size n of the sequence, and of the arbitrary choices of blocks size r and mixing range m, *c.f.* Equation (5.6) in Chapter 5. Referring to Definition 2.3.2, this implies that

$$\left| \mathbb{E} \sum_{1 \leq k \leq n} \mathbb{I}_{\{X_k > u\}} - \mathbb{E}\left[\mathrm{CP}\left(\theta_r \boldsymbol{\mu}([0, np]), \nu_r\right)\right] \right| \leq n\epsilon. \tag{4.5}$$

Because of the general character of this result, without any assumption made on the original sequence $(X_i)_{1 \leq i \leq n}$, the parameters of the empirical compound Poisson approximation in the paper could not be calculated. But thanks to the results of Section 4.4, more information are available on the asymptotic model. This is why three things will be calculated:

1. the total count of exceedances in the sample, namely $\sum_{1 \leq k \leq n} \mathbb{I}_{\{X_k > u\}}$;
2. the expectation of the asymptotic compound Poisson distribution for the total number of counts, namely $\mathbb{E}\left[\mathrm{CP}\left(\theta_{|X|} x^{-2\kappa} \boldsymbol{\mu}(]0, 1]), \nu\right)\right]$;
3. the bound on the empirical approximation (the available bound on the asymptotic model being larger, for triangular inequality reasons).

We shall finally look at $|1. - 2.|$ on the one hand, and 3. on the other hand, and finding out that $3. \gg |1. - 2.|$ would be a hint that the bound might be slack.

4.5.4 Figures

The expression of the bound in Barbour *et al.* [7] for m-dependent processes, with block size $r = \sqrt{mn}$ and exceedance probability $p = \mathbb{P}(X_1 > u)$ is

$$\epsilon = p\left(6\sqrt{mn} + m\left(e^{mp} + \frac{1.65}{\sqrt{1 - mp}}\right)\right)$$

As an example, this is computed for 2-dependent observations. Indeed, the bound is an increasing function of m, so if the bound is slack for $m = 2$, then it will be slacker

for larger values. The value of p plugged in is the estimator $\widehat{p} = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{\{X_k > u\}}$. Let us point out that this is a very rough approach. Indeed, $\widehat{p}$ is only computed for the simulated sample, and no confidence interval is calculated, as would be commanded by any decent statistical approach. The aim here is just to get a rough idea of what is going on. If one wants to estimate p properly, then Aebi *et al.* [1] provide a good method to estimate the distribution of perpetuities, using bootstrap techniques. Quantities 1., 2. and 3. are then computed for nine different threshold values, logarithmically spaced towards the end of the distribution tail, giving Table 4.1. Noting $X_{\max} = \max\{X_k, 1 \leq k \leq 2248\}$, the nine thresholds are $\forall i \in [\![1, 9]\!], u_i = X_{\max}(1-\frac{1}{2^i})$.

Threshold index	1	2	3	7	8	9
Empirical count (1.)	21	1	1	1	1	1
Asymptotic expectation (2.)	453.3	26.35	8.936	3.702	3.594	3.552
Bound (3.)	1449	145.0	56.40	25.61	24.94	24.68

Table 4.1: *Values of the bound in [7] for different thresholds.*

We are therefore in a $3. \gg |1. - 2.|$ situation, so our heuristic analysis hints at a slack bound. This is probably due to the absence of assumptions on the sequence, which prevents us as well from accessing an explicit expression of the empirical approximation. Part of the motivation for the present work is to make use of the particular dependence structure available in a perpetuity.

Chapter 5

Case of finite, $\{0,1\}$-valued, first-order Markov chains

This is now the first chapter of this work where a $\{0,1\}$-valued process is linked to an empirical compound Poisson measure. Unfortunately, even though a first-order stochastic recursion forms a first-order Markov chain, its exceedance process does not. More precisely, the latter is the observed part of a hidden Markov chain. This means we shall have to wait a few more chapters to be able to get an overall approximation of the exceedances by a compound Poisson measure. Still, before moving on to the more complex approximations of $\{0,1\}$-valued d-th order Markov chains and m-dependent processes, this constitutes a first result where the spirit of the method can be explained. In this section, the approach in Barbour *et al.* [7] is adapted to the special case of finite, $\{0,1\}$-valued, first-order Markov chains. Not only will the bound in Barbour *et al.* [7] be tightened, but explicit expressions for the empirical model will be obtained, opening the door to practical applications. The

main result is Theorem 5.2.1. Even though our exceedance process does not form a Markov chain, nothing prevents us from giving pertinent properties to the Markov chain we are going to study. One can very well decide to only look at exceedances, irrespective of the properties of the underlying process. In such a perspective, one can choose to model them directly as a Markov chain and forget all assumptions of stochastic recurrence.

5.1 General idea of the method

Throughout this chapter we shall assume that $(I_i)_{1\leq i\leq n}$ is a $\{0,1\}$-valued, first-order, strictly stationary, time-homogeneous Markov chain such that $p = \mathbb{P}(I_1 = 1) \in]0,1[$, $p_1 = \mathbb{P}(I_2 = 1 | I_1 = 0) \in]0,1[$ and also $\mathbb{P}(I_2 = 1 | I_1 = 1) \in]0,1[$. These properties make the state space $\{0,1\}$ small, and hence petite for $(I_i)_{1\leq i\leq n}$. In addition, they also clearly make $(I_i)_{1\leq i\leq n}$ a φ-irreducible and aperiodic chain. Consequently, Theorem 16.2.2 in Meyn and Tweedie [51] tells us that our Markov chain is uniformly ergodic. In addition, let us denote

$$N_n(\bullet) = \sum_{i=1}^{n} I_i \delta_{ip}(\bullet) \qquad \text{and} \qquad \forall i \in \left\{1, \ldots, \left\lfloor \frac{n}{r} \right\rfloor\right\}, T_{r,i} = \sum_{k=(i-1)r+1}^{ir} I_k.$$

The idea behind the method is to use triangular inequalities to bound distances between judiciously chosen auxiliary point processes. Recall from Chapter 4 that Bernstein blocks create the following point process on $[0, np]$, *c.f.* Equation (4.3):

$$\sum_{i=1}^{\lfloor n/r \rfloor} T_{r,i} \delta_{irp}(\bullet)$$

and heuristically, under suitable mixing conditions, the sequence $(T_{r,i})_{1\leq i\leq \lfloor n/r \rfloor}$ of counts of 1's per block should be close to i.i.d. if the block size r is sufficiently large. We shall therefore use an i.i.d. sequence $(\widehat{T}_{r,i})_{1\leq i\leq \lfloor n/r \rfloor}$ such that $\widehat{T}_{r,1} \overset{d}{=} T_{r,1}$. Xia [68] gives a bound on how far the related point process is from a certain discrete empirical compound Poisson measure.

Now, let us inquire about the distance between $\sum T_{r,i}\delta_{irp}(\bullet)$ and $\sum \widehat{T}_{r,i}\delta_{irp}(\bullet)$. Here is where we take the dependence structure of our original sequence into account. If we have reasons to believe that there is a significant dependence in $(I_i)_{1\leq i\leq n}$ over a range of about m lags, then let us form the following quantities

$$\forall i \in \left\{1,\ldots,\left\lfloor \frac{n}{r} \right\rfloor\right\}, \quad T_{r,i}^{(m)} = \sum_{k=(i-1)r+1}^{ir-m} I_k,$$

In other words, $T_{r,i}^{(m)}$ is the number of 1's in the i^{th} block deprived from its m last elements. That way, if $(I_i)_{1\leq i\leq n}$ was m-dependent, then $T_{r,i}^{(m)}$ and $T_{r,j}^{(m)}$ would be independent for $i \neq j$. For large values of m, under suitable mixing conditions, it is plausible that $T_{r,i}^{(m)}$ and $T_{r,j}^{(m)}$ would be approximately independent for i and j not too close. Hence, as before, we shall use an i.i.d. sequence $(\widehat{T}_{r,i}^{(m)})_{1\leq i\leq \lfloor n/r\rfloor}$, where $\widehat{T}_{r,1}^{(m)} \stackrel{d}{=} T_{r,1}^{(m)}$. More precisely, let us choose our sequences $(\widehat{T}_{r,i})_{1\leq i\leq \lfloor n/r\rfloor}$ and $(\widehat{T}_{r,i}^{(m)})_{1\leq i\leq \lfloor n/r\rfloor}$ such that $\left((\widehat{T}_{r,i}, \widehat{T}_{r,i}^{(m)})\right)_{1\leq i\leq \lfloor n/r\rfloor}$ are i.i.d. replicas of $(T_{r,1}, T_{r,1}^{(m)})$. As in Barbour *et al.* [7], the compound Poisson parameters are $\theta_r = \frac{1}{rp}\,\mathbb{P}\left(T_{r,1}\geq 1\right)$ and $\nu_r = \mathcal{L}\left(T_{r,1} \middle| T_{r,1}\geq 1\right)$, and we shall calculate them for our present situation in the next section. The triangular inequality used is the following.

$$\begin{aligned}
& d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\
\leq\ & d_W\left(\mathcal{L}\big(N_n(\bullet)\big), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right)\right) && (5.1)\\
+\ & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) && (5.2)\\
+\ & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) && (5.3)\\
+\ & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right)\right) && (5.4)\\
+\ & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\right). && (5.5)
\end{aligned}$$

In the next section, we shall calculate bounds for each of these quantities. A bound on (5.5) is derived in *ibid.* [7] from Xia [68], where Stein's method is applied to the

compound Poisson distribution, and there is no obvious improvement to be gained by knowing the Markovian structure of $(I_i)_{1\leq i\leq n}$. A bound on (5.1) is easily found via distortion considerations, *c.f.* Equation (2.13) in *ibid.* [7]. Indeed, from the point processes $N_n(\bullet)$ to $\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)$, the random, $\{0,1\}$-valued weights are simply gathered at the end of each block, which implies that the average distortion is no bigger than the width of one block, *i.e.* no larger than rp. The last, incomplete block comprising the points with indices in $\{r\lfloor n/r\rfloor + r,\ldots,n\}$ is omitted, an hence this introduces an error of at most rp. Overall, we end up with a bound of $2rp$ on (5.1). Now, for the rest of the right-hand side in the triangular inequality, namely (5.2), (5.3) and (5.4), the Markovian nature of $(I_i)_{1\leq i\leq n}$ enables us to calculate a new, tighter bound than the one in Barbour *et al.* [7], which is

$$\begin{aligned} & d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r\boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\ & \quad \leq \ 4rp + \frac{2n}{r}mp + \frac{n}{r}\beta(m) + q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right), \end{aligned} \tag{5.6}$$

where $q_r = \mathbb{P}\left(T_{r,1}\geq 1\right)$, and β is the absolute regularity coefficient of $(I_t)_{1\leq t\leq n}$.

5.2 Link between the first-order Markov process and its related compound Poisson measure

Given that we are dealing with a stationary, time-homogeneous Markov chain, let us introduce the quantity $p_k = \mathbb{P}\left(I_t = 1 \middle| I_{t-k} = 0\right)$, which does not depend on t. With this tool, let us return to the bound. Lemmas 5.2.1 to 5.2.3 are original, and lead to Theorem 5.2.1, which gives an explicit approximation for the occurrence of 1 in a $\{0,1\}$-valued Markov chain.

Lemma 5.2.1 *With the notation* $p_k = \mathbb{P}\left(I_t = 1 \,\middle|\, I_{t-k} = 0\right)$ *and* $p = \mathbb{P}\left(I_1 = 1\right)$, *we*

have

$$d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\Big)\right) \leq 1-(1-p)(1-p_1)^{\lfloor \frac{n}{r}\rfloor(m-1)}(1-p_{r-m+1})^{\lfloor \frac{n}{r}\rfloor-1}.$$

Remark. Heuristically, let us see when this bound is small. The type of $\{0,1\}$-valued variables we are eventually looking at in this work are exceedances above high thresholds. This means the exceedance probability p and transition probabilities p_k involved are small. They also depend on the length n of the sequence, as the threshold $u = u_n$ does. Indeed, the probabilities will vanish as n increases. Hence, for a fixed n, $1-p$ and $1-p_{r-m+1}$ are close to 1, and if m is not so large that $(1-p_1)^m$ becomes small, then the bound is small.

Proof of Lemma 5.2.1. By Corollary 2.3.1, we have

$$\begin{aligned}
&d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\Big)\right)\\
&\leq d_{TV}\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\Big)\right)\\
&\leq \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet) \neq \sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\\
&= 1-\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet) = \sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\\
&= 1-\mathbb{P}\left(\forall i\in[\![1,\lfloor n/r\rfloor]\!], T_{r,i} = T_{r,i}^{(m)}\right)\\
&= 1-\mathbb{P}\left(\forall i\in[\![1,\lfloor n/r\rfloor]\!], \sum_{k=ir-m+1}^{ir} I_k = 0\right)\\
&= 1-\mathbb{P}\left(\forall k\in\overline{B}, I_k = 0\right)
\end{aligned}$$

where $\overline{B} = \bigcup_{i=1}^{\lfloor n/r\rfloor}\{ir-m+1,\ldots,ir\}$.

Now, the set $\left\{\forall k\in\overline{B}, I_k = 0\right\}$ describes the situation in which all the elements of the Markov chain $\left(I_k\right)_{1\leq k\leq n}$ are equal to 0 on the last m spots of each block. To

obtain such a configuration, we first need to have 0 on the "first" element of $\overline{B}$, which happens with probability $1-p$, then 0 on the next one, which adds a multiplicative factor of $(1-p_1)$ to the probability, and so on until the end of the first block. Then, we need to have the first of the m last elements in the second block equal to zero, which adds a multiplicative factor of $(1-p_{r-m+1})$ to the probability. Indeed, we make a jump of $r-m+1$ lags since the last event. Carrying on like this, and given that there are $\lfloor \frac{n}{r} \rfloor$ blocks which imply $\lfloor \frac{n}{r} \rfloor - 1$ jumps of $r-m+1$ lags, and m elements at the end of each block, which imply $\lfloor \frac{n}{r} \rfloor (m-1)$ jumps of one lag, we end up with

$$\mathbb{P}\left(\forall\, k \in \overline{B}, I_k = 0\right) = (1-p)(1-p_1)^{\lfloor \frac{n}{r} \rfloor (m-1)} (1-p_{r-m+1})^{\lfloor \frac{n}{r} \rfloor - 1}.$$

Hence the result. □

Let us now use a similar line of argument to bound (5.4).

Lemma 5.2.2 *With the notation of Lemma 5.2.1, we have*

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right) \\ \leq\ & 1 - \left((1-p)(1-p_1)^{(m-1)}\right)^{\lfloor \frac{n}{r} \rfloor}. \end{aligned}$$

Remark. Again, if p is small and m not too large, then this bound is small.

Proof of Lemma 5.2.2. Again by Corollary 2.3.1, we have

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right) \\ \leq\ & d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right) \\ \leq\ & \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet) \neq \sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right) \\ =\ & 1 - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet) = \sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right) \\ =\ & 1 - \mathbb{P}\left(\forall\, i \in [\![1, \lfloor n/r \rfloor]\!], \widehat{T}_{r,i} = \widehat{T}^{(m)}_{r,i}\right) \\ =\ & 1 - \left(\mathbb{P}\left(\widehat{T}_{r,1} = \widehat{T}^{(m)}_{r,1}\right)\right)^{\lfloor n/r \rfloor}, \end{aligned} \tag{5.7}$$

because $\left((\widehat{T}_{r,i}, \widehat{T}_{r,i}^{(m)})\right)_{1\leq i\leq \lfloor n/r\rfloor}$ are i.i.d. Now, given that $(\widehat{T}_{r,1}, \widehat{T}_{r,1}^{(m)})$ is an independent replica of $(T_{r,1}, T_{r,1}^{(m)})$, we can substitute in Equation (5.6), and obtain

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) \\ \leq\ & 1-\left(\mathbb{P}\left(T_{r,1}=T_{r,1}^{(m)}\right)\right)^{\lfloor n/r\rfloor} \\ =\ & 1-\left(\mathbb{P}\left(\sum_{k=r-m+1}^{r} I_k=0\right)\right)^{\lfloor n/r\rfloor} \\ =\ & 1-\left(\mathbb{P}\left(\forall\, k\in[\![r-m+1,r]\!], I_k=0\right)\right)^{\lfloor n/r\rfloor}. \end{aligned}$$

Now, in order to obtain zeros for all I_k between $k=r-m+1$ and $k=r$, we have to start with one for the first, which happens with probability $1-p$. Then, the $m-1$ last elements have to be put to zero as well, each adding a multiplicative factor $(1-p_1)$ to the probability. Consequently, we end up with

$$\mathbb{P}\left(\forall\, k\in[\![r-m+1,r]\!], I_k=0\right) = (1-p)(1-p_1)^{(m-1)}.$$

Hence the result. □

Now let us tackle Term (5.3), concerning the approximation by i.i.d. blocks of the process with m elements removed at the end of each block. This is where the mixing structure of the process becomes apparent.

Lemma 5.2.3 *Still with the same notation,*

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) \leq \frac{2n}{r}(1-p)|p_{m+1}-p|\,.$$

Remark. This bound is small if $\frac{n}{r}$ is not too large and if m is large enough for $|p_{m+1}-p|$ to be small. Indeed, the latter is in fact $|p_{m+1}-p| = |\mathbb{P}\left(I_{m+1}=1|I_1=0\right) - \mathbb{P}\left(I_{m+1}=1\right)|$. Now, given that (I_t) is uniformly ergodic, Theorem 16.2.1 in Meyn and Tweedie [51] shows that it is also geometrically ergodic. Hence, there exist $\rho\in]0,1[$ and $C>0$ such that $|p_{m+1}-p|\leq C\rho^m$. It is obvious from this expression that the bound is small for large values of m.

Proof of Lemma 5.2.3. We know from Barbour *et al.* [7] that, using Lindeberg's method of decomposition (*c.f.* Lindeberg [48]), and then Berbee's Lemma (*c.f.* Berbee [14]),

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) \leq \frac{n}{r}\beta(m).$$

Recall from Section 2.4.2 that β is the mixing coefficient of absolute regularity defined as

$$\beta(l) = \max_{1\leq k\leq n} \mathbb{E} \sup_{H\in\mathcal{F}_{k+l+1,n}} \left|\mathbb{P}(H|\mathcal{F}_{1,k}) - \mathbb{P}(H)\right|$$

where $\forall\,(s,t)\in\{1,\ldots,n\}^2$, $s\leq t$, $\mathcal{F}_{s,t}=\sigma\big(I_k,\ s\leq k\leq t\big)$. Let us calculate β in our special Markovian case. We know from Proposition 1 in Davydov [26] that, for a time-homogeneous Markov chain $(X_t)_{t\in\mathbb{N}^*}$ on a state space $(S,\mathcal{S})$, we have

$$\beta(m) \;=\; \sup_{t\in\mathbb{N}} \int_S \mathbb{P}(X_t\in dx) d_{TV}\Big(\mathcal{L}(X_{m+1}|X_1=x), \mathcal{L}(X_{t+m})\Big).$$

Here, we are dealing with a strictly stationary process, hence

$$\beta(m) \;=\; \int_S \mathbb{P}(X_1\in dx) d_{TV}\Big(\mathcal{L}(X_{m+1}|X_1=x), \mathcal{L}(X_{m+1})\Big).$$

The state space here is $S=\{0,1\}$, therefore

$$\beta(m) \;=\; \sum_{i=0}^{1} \mathbb{P}(I_1=i) d_{TV}\Big(\mathcal{L}(I_{m+1}|I_1=i), \mathcal{L}(I_{m+1})\Big), \tag{5.8}$$

and

$$d_{TV}\Big(\mathcal{L}(I_{m+1}|I_1=i), \mathcal{L}(I_{m+1})\Big) \;=\; \max_{j\in\{0,1\}} \big|\mathbb{P}(I_{m+1}=j|I_1=i) - \mathbb{P}(I_{m+1}=j)\big|.$$

Given that $\mathbb{P}(I_1=1)=1-\mathbb{P}(I_1=0)$, we have

$$\max_{j\in\{0,1\}} \big|\mathbb{P}(I_{m+1}=j|I_1=i) - \mathbb{P}(I_{m+1}=j)\big| \;=\; \big|\mathbb{P}(I_{m+1}=1|I_1=i) - \mathbb{P}(I_{m+1}=1)\big|,$$

and after substitution in (5.8), we get

$$\begin{aligned}\beta(m) \;&=\; \sum_{i=0}^{1} \mathbb{P}(I_1=i)\big|\mathbb{P}(I_{m+1}=1|I_1=i) - \mathbb{P}(I_{m+1}=1)\big| \\ &=\; p|\mathbb{P}(I_{m+1}=1|I_1=1) - p| + (1-p)|p_{m+1} - p|.\end{aligned}$$

Now, after conditioning on the value of X_1, we clearly have $p = p\mathbb{P}(I_{m+1} = 1|I_1 = 1) + (1-p)p_{m+1}$. Consequently,

$$\begin{aligned} p\mathbb{P}(I_{m+1} = 1|I_1 = 1) - p^2 &= p - p^2 - (1-p)p_{m+1} \\ &= (1-p)(1-p_{m+1}). \end{aligned}$$

We therefore obtain the following value for the absolute regularity coefficient, in the case of an infinite sequence:

$$\beta(m) = 2(1-p)|p_{m+1} - p|$$

Lemma 2.4.1 indicates that this bounds the β-mixing coefficient associated with our finite sequence, which completes the proof. □

Such calculations are related to the ones developed in Athreya and Pantula [3] to prove that some families of Markov chains are strongly mixing. Adding the bounds on expressions (5.1) to (5.5), this finally enables us to obtain a bound on the whole compound Poisson approximation, with parameters $\theta_r = \frac{1}{rp}\,\mathbb{P}\,(T_{r,1} \geq 1)$ and $\nu_r = \mathcal{L}\,(T_{r,1}|\,T_{r,1} \geq 1)$. We shall investigate their explicit form in the next section.

Theorem 5.2.1 (Empirical approximation) *Still with the same notation,*

$$\begin{aligned} d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \leq\ & 4rp + \frac{2n}{r}(1-p)|p_{m+1} - p| \\ & + 1 - \Big((1-p)(1-p_1)^{(m-1)}\Big)^{\lfloor \frac{n}{r} \rfloor} \\ & + 1 - (1-p)(1-p_1)^{\lfloor \frac{n}{r} \rfloor (m-1)}(1-p_{r-m+1})^{\lfloor \frac{n}{r} \rfloor - 1} \\ & + q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right), \end{aligned}$$

where $q_r = 1-(1-p)(1-p_1)^{(r-1)}$.

Proof. It is now just a matter of bounding all the elements (5.1) to (5.5) in the right-hand side of the big triangular inequality of Section 5.1. The bound for (5.1)

$$d_W\left(\mathcal{L}\big(N_n(\bullet)\big), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} T_{r,i}\delta_{irp}(\bullet)\right)\right) \leq 2rp$$

is provided in [7], as well as the bound for (5.5)

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}_{r,i}\delta_{irp}(\bullet)\right),\mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right) \;\leq\; 2rp+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),$$

where $q_r=\mathbb{P}\left(T_{r,1}\geq 1\right)$, directly derived from an earlier result on the Poisson approximation using Stein's method by Xia [68]. Now, with an argument analogous to these of Lemmas 5.2.1 and 5.2.2 , noting that $q_r=1-\mathbb{P}\left(\{I_1=0,\ldots,I_r=0\}\right)$, we get

$$q_r=1-(1-p)(1-p_1)^{(r-1)}. \tag{5.9}$$

Lemmas 5.2.1, 5.2.2, and 5.2.3 complete the proof. □

5.3 Empirical compound Poisson parameters

The approximating model is $\mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)$, with $\theta_r=\frac{1}{rp}\mathbb{P}\left(T_{r,1}\geq 1\right)$ and $\nu_r=\mathcal{L}\left(T_{r,1}|\,T_{r,1}\geq 1\right)$. We can immediately derive from (5.9) in the proof of Theorem 5.2.1 that

$$\boxed{\theta_r=\frac{1}{rp}\left(1-(1-p)(1-p_1)^{(r-1)}\right)}.$$

Let us now focus on the empirical compounding distribution ν_r. If T is a random variable with distribution ν_r, and if $k\in\mathbb{N}$, our aim is to find $\mathbb{P}\left(T=k\right)$. We have $\nu_r=\mathcal{L}\left(T_{r,1}|\,T_{r,1}\geq 1\right)$, hence

$$\mathbb{P}\left(T=k\right)=\mathbb{P}\left(T_{r,1}=k|\,T_{r,1}\geq 1\right)=\mathbb{P}\left(T_{r,1}=k,T_{r,1}\geq 1\right)/\,\mathbb{P}\left(T_{r,1}\geq 1\right).$$

We already know $\mathbb{P}\left(T_{r,1}\geq 1\right)$ from (5.9) in the proof of Theorem 5.2.1. Clearly, $\mathbb{P}\left(T=0\right)=0$ and $\mathbb{P}\left(T=k\right)=0$ for $k\geq r+1$. We can therefore assume from now on that $k\in[\![1,r]\!]$. With that assumption, $\mathbb{P}\left(T_{r,1}=k,T_{r,1}\geq 1\right)=\mathbb{P}\left(T_{r,1}=k\right)=\mathbb{P}\left(\sum_{s=1}^{r}I_s=k\right)$.

It all boils down to the following problem. In a Markovian sequence of size r, with two possible states 0 and 1, if $k\in[\![1,r]\!]$, what is the probability of getting the output 1 exactly k times?

The answer to this can be derived from Goodman [38]. The main result there concerns the distribution of the frequency matrix of pairs of consecutive outcomes.

Theorem 5.3.1 (Goodman) *Let $Y_1, \ldots, Y_N$ be a finite sequence from a Markov chain with q different states, constant transition matrix with all coefficients in $]0,1[$. Calling $F_{i,j}$ the frequency of the sequence (i,j), we have*

$$F_{i,j} = \sum_{s=1}^{N-1} \mathbb{I}_{\left\{(Y_s, Y_{s+1})=(i,j)\right\}}.$$

Let us now call $F = [F_{i,j}]_{1\leq i,j\leq q}$ the random matrix of frequencies, and denote $f = [f_{i,j}]_{1\leq i,j\leq q}$ its realisation at a given point $\omega \in \Omega$, so that $F(\omega) = f$. Let $F_{i\cdot} = \sum_{j=1}^{q} F_{i,j}$, $f_{i\cdot} = \sum_{j=1}^{q} f_{i,j}$ and let $P = [p_{i,j}]_{1\leq i,j\leq q}$ be the transition matrix for the Markov chain. Let $T(F,w,v)$ be the (w,v)th cofactor of the matrix $\widehat{M} = [\delta_{i,j} - F_{i,j}/F_{i\cdot}]_{1\leq i,j\leq q}$. Then

$$\mathbb{P}\left((F, Y_N) = (f,w) \middle| Y_1 = v\right) = T(f,w,v)\, \frac{\prod_{i=1}^{q} f_{i\cdot}!}{\prod_{i=1}^{q}\prod_{j=1}^{q} f_{i,j}!} \prod_{i=1}^{q}\prod_{j=1}^{q} p_{i,j}^{f_{i,j}}.$$

This result is derived in [38] from the BEST theorem, using combinatorial arguments for oriented linear graphs. Let us now apply it to our situation. We put $Y_s = \mathbb{I}_{\{X_s \geq u\}}$, $N = r$ and $q = 2$, hence we can adopt the notations

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix},\ f(a,b,c,d) = \begin{pmatrix} a & b \\ c & d \end{pmatrix},\ \text{and } \mathrm{Co}(\widehat{M}) = \begin{pmatrix} \frac{c}{c+d} & \frac{c}{c+d} \\ \frac{b}{a+b} & \frac{b}{a+b} \end{pmatrix},$$

where f is any matrix representing the possible counts of frequencies, *i.e.* any value of the random matrix F. For the sake of convenience, we have denoted $f_{1,1} = a$, $f_{1,2} = b$, $f_{2,1} = c$ and $f_{2,2} = d$. This implies that $T(f,0,0) = T(f,0,1) = \frac{c}{c+d}$, and $T(f,1,0) = T(f,1,1) = \frac{b}{a+b}$. In order to find $\mathbb{P}(T=k)$, we saw earlier that we just had to find $\mathbb{P}\left(\sum_{s=1}^{r} \mathbb{I}_{\{Y_s=1\}} = k\right)$. Now $\sum_{s=1}^{r} \mathbb{I}_{\{Y_s=1\}} = F_{1,0} + F_{1,1} + \mathbb{I}_{\{Y_r=1\}}$. Hence

$$\begin{aligned} &\mathbb{P}\left(\sum_{s=1}^{r} \mathbb{I}_{\{Y_s=1\}} = k\right) \\ &= \mathbb{P}\left(F_{1,0} + F_{1,1} + \mathbb{I}_{\{Y_r=1\}}\right) \\ &= p\,\mathbb{P}\left(F_{1,0}+F_{1,1}=k-1 \,\middle|\, Y_r=1\right) + (1-p)\,\mathbb{P}\left(F_{1,0}+F_{1,1}=k \,\middle|\, Y_r=0\right). \end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathbb{P}\left(F_{1,0}+F_{1,1}=k \,\middle|\, Y_r=0\right) \\
&= \sum_{\substack{c+d=k \\ a+b+c+d=r-1}} \mathbb{P}\left(F_{1,0}+F_{1,1}=k \,\middle|\, Y_r=0, F=f(a,b,c,d)\right) \mathbb{P}\left(F=f \,\middle|\, Y_r=0\right) \\
&= \sum_{\substack{c+d=k \\ a+b+c+d=r-1}} \mathbb{P}\left(F=f(a,b,c,d) \,\middle|\, Y_r=0\right).
\end{aligned}$$

We limit the range of the sum of all terms of f to $r-1$, because that is the total number of 2-tuples of consecutive terms in a sequence of length r. Now, using the result from Theorem 5.3.1, we get

$$\mathbb{P}\left(F_{1,0}+F_{1,1}=k \,\middle|\, Y_r=0\right) = \sum_{\substack{c+d=k \\ a+b=r-k-1}} \frac{1}{1-p} \cdot \frac{c}{c+d} \cdot \frac{(a+b)!\,(c+d)!}{a!\,b!\,c!\,d!} \alpha^a \beta^d (1-\alpha)^b (1-\beta)^c.$$

After a similar argument for $\mathbb{P}\left(F_{1,0}+F_{1,1}=k-1 \,\middle|\, Y_r=1\right)$, we get the following expression.

$$\begin{aligned}
&\mathbb{P}\left(\sum_{s=1}^{r} \mathbb{I}_{\{Y_s=1\}}=k\right) \\
&= \sum_{\substack{c+d=k-1 \\ a+b=r-k}} \frac{b}{a+b} \cdot \frac{(a+b)!\,(c+d)!}{a!\,b!\,c!\,d!} \alpha^a \beta^d (1-\alpha)^b (1-\beta)^c \\
&\quad + \sum_{\substack{c+d=k \\ a+b=r-k-1}} \frac{c}{c+d} \cdot \frac{(a+b)!\,(c+d)!}{a!\,b!\,c!\,d!} \alpha^a \beta^d (1-\alpha)^b (1-\beta)^c \\
&= \sum_{c=0}^{k-1} \sum_{a=0}^{r-k} \frac{r-k-a}{r-k} \cdot \frac{(r-k)!\,(k-1)!}{a!\,(r-k-a)!\,c!\,(k-1-c)!} \alpha^a \beta^{k-1-c} (1-\alpha)^{r-k-a} (1-\beta)^c \\
&\quad + \sum_{c=0}^{k} \sum_{a=0}^{r-k-1} \frac{c}{k} \cdot \frac{(r-k-1)!\,(k)!}{a!\,(r-k-1-a)!\,c!\,(k-c)!} \alpha^a \beta^{k-c} (1-\alpha)^{r-k-1-a} (1-\beta)^c \\
&= \sum_{c=0}^{k-1} \sum_{a=0}^{r-k-1} \frac{(r-k-1)!\,(k-1)!}{a!\,(r-k-a-1)!\,c!\,(k-1-c)!} \alpha^a \beta^{k-1-c} (1-\alpha)^{r-k-a} (1-\beta)^c \\
&\quad + \sum_{c=1}^{k} \sum_{a=0}^{r-k-1} \frac{(r-k-1)!\,(k-1)!}{a!\,(r-k-1-a)!\,(c-1)!\,(k-c)!} \alpha^a \beta^{k-c} (1-\alpha)^{r-k-1-a} (1-\beta)^c.
\end{aligned}$$

Let us now factorise the terms together.

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{s=1}^{r}\mathbb{I}_{\{Y_s=1\}}=k\right)\\
&= \sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!}{a\,!\,(r-k-a-1)\,!}\,\alpha^a\beta^{k-1}(1-\alpha)^{r-k-a}\\
&\quad+\sum_{c=1}^{k-1}\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!\,(k-1)\,!}{a\,!\,(r-k-a-1)\,!\,c\,!\,(k-1-c)\,!}\,\alpha^a\beta^{k-1-c}(1-\alpha)^{r-k-a}(1-\beta)^c\\
&\quad+\sum_{c=1}^{k-1}\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!\,(k-1)\,!}{a\,!\,(r-k-1-a)\,!\,(c-1)\,!\,(k-c)\,!}\,\alpha^a\beta^{k-c}(1-\alpha)^{r-k-1-a}(1-\beta)^c\\
&\quad+\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!}{a\,!\,(r-k-1-a)\,!}\,\alpha^a(1-\alpha)^{r-k-1-a}(1-\beta)^k\\
&= \sum_{c=1}^{k-1}\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!\,(k-1)\,!\alpha^a\beta^{k-1-c}(1-\alpha)^{r-k-a-1}(1-\beta)^c}{a\,!\,(r-k-a-1)\,!\,(c-1)\,!\,(k-1-c)\,!}\left(\frac{1-\alpha}{c}+\frac{\beta}{k-c}\right)\\
&\quad+\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!}{a\,!\,(r-k-1-a)\,!}\,\alpha^a(1-\alpha)^{r-k-1-a}(1-\beta)^k\\
&\quad+\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!}{a\,!\,(r-k-a-1)\,!}\,\alpha^a\beta^{k-1}(1-\alpha)^{r-k-a}\\
&= \sum_{c=1}^{k-1}\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!\,(k-1)\,!\alpha^a\beta^{k-1-c}(1-\alpha)^{r-k-a-1}(1-\beta)^c}{a\,!\,(r-k-a-1)\,!\,(c-1)\,!\,(k-1-c)\,!}\left(\frac{1-\alpha}{c}+\frac{\beta}{k-c}\right)\\
&\quad+\sum_{a=0}^{r-k-1}\frac{(r-k-1)\,!}{a\,!\,(r-k-1-a)\,!}\,\alpha^a(1-\alpha)^{r-k-1-a}\Big((1-\beta)^k+\beta^{k-1}(1-\alpha)\Big)\\
&= \sum_{c=1}^{k-1}\sum_{a=0}^{r-k-1}\frac{1}{k}\mathrm{C}^a_{r-k-1}\mathrm{C}^c_k\alpha^a\beta^{k-1-c}(1-\alpha)^{r-k-a-1}(1-\beta)^c\Big((1-\alpha)(k-c)+c\beta\Big)\\
&\quad+\sum_{a=0}^{r-k-1}\mathrm{C}^a_{r-k-1}\alpha^a(1-\alpha)^{r-k-1-a}\Big((1-\beta)^k+\beta^{k-1}(1-\alpha)\Big)\\
&= \sum_{a=0}^{r-k-1}\mathrm{C}^a_{r-k-1}\alpha^a(1-\alpha)^{r-k-1-a}\\
&\quad\times\left((1-\beta)^k+\beta^{k-1}(1-\alpha)+\sum_{c=1}^{k-1}\frac{1}{k}\mathrm{C}^c_k\beta^{k-1-c}(1-\beta)^c\Big((1-\alpha)(k-c)+c\beta\Big)\right)\\
&= \sum_{a=0}^{r-k-1}\mathrm{C}^a_{r-k-1}\alpha^a(1-\alpha)^{r-k-1-a}\left(\sum_{c=0}^{k}\frac{1}{k}\mathrm{C}^c_k\beta^{k-1-c}(1-\beta)^c\Big((1-\alpha)(k-c)+c\beta\Big)\right)\\
&= \sum_{a=0}^{r-k-1}\mathrm{C}^a_{r-k-1}\alpha^a(1-\alpha)^{r-k-1-a}\left(\sum_{c=0}^{k}\frac{1}{k}\mathrm{C}^c_k\beta^{k-c}(1-\beta)^c\Big(c+\frac{1-\alpha}{\beta}(k-c)\Big)\right),
\end{aligned}
$$

where $\mathrm{C}_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$. Therefore, we finally obtain

$$\mathcal{L}(T)=\nu_r \Longrightarrow \forall\, k\in[\![1,r]\!],$$
$$\mathbb{P}\,(T=k) = \frac{1}{k}\sum_{a=0}^{r-k-1} \mathrm{C}_{r-k-1}^{a}\alpha^a(1-\alpha)^{r-k-1-a}\left(\sum_{c=0}^{k}\mathrm{C}_k^c\beta^{k-c}(1-\beta)^c\left(c+\frac{1-\alpha}{\beta}(k-c)\right)\right).$$

Chapter 6

Case of finite, $\{0,1\}$-valued, d-th order Markov chains

For multivariate exceedances, it might be the case that a d-th order Markov chain be more adequate to describe the exceedance point process. This chapter deals with the related approximation. This new viewpoint is especially fruitful because it prepares the grounds to deal with m-dependent processes, even though it constitutes a more general framework. But yet again, one might choose to ignore the structure of the underlying process, only look at its exceedances, and directly model them as a d-th order Markov chain.

In this new framework, throughout this chapter we shall assume that $\left(I_i\right)_{1\leq i\leq n}$ is a $\{0,1\}$-valued, d-th order, strictly stationary, time-homogeneous Markov chain. Let us adopt the appropriate notation. If $\forall\, i \in [\![1,d]\!], k_i \in \mathbb{N}$, and $\boldsymbol{k} \in \mathbb{N}^d$ such that

$\boldsymbol{k} = (k_1, \ldots, k_d)$, let

$$p_{\boldsymbol{k}} \quad = \quad \mathbb{P}\left(I_t = 1 \,|\, I_{t-k_1} = 0, \ldots, I_{t-k_1-\cdots-k_d} = 0\right).$$

Now, let $\boldsymbol{v}_i \in \mathbb{N}^d$ such that $\boldsymbol{v}_i = \left(\mathbb{I}_{\{i\geq 1\}}, \ldots, \mathbb{I}_{\{i\geq d\}}\right)$, and let $\boldsymbol{w}_i \in \mathbb{N}^d$ such that

$$\boldsymbol{w}_i = (1, \ldots, 1, r-m+1, 1, \ldots, 1).$$
$$\uparrow$$
$$i\text{-th position}$$

We shall assume that for all $\boldsymbol{k} \in \mathbb{N}^d$, $p_{\boldsymbol{k}}$ is in $]0,1[$. Consequently, any d-th order transition happens with a non-zero probability. This property makes the state space $\{0,1\}$ small, and hence petite for $\left(I_i\right)_{1\leq i\leq n}$. In addition, it also clearly makes $\left(I_i\right)_{1\leq i\leq n}$ a ψ-irreducible and aperiodic chain. Consequently, Theorem 16.2.2 in Meyn and Tweedie [51] tells us that our Markov chain is uniformly ergodic. As in the last chapter, let us denote

$$N_n(\bullet) = \sum_{i=1}^{n} I_i \delta_{ip}(\bullet) \qquad \text{and} \qquad \forall i \in \left\{1, \ldots, \left\lfloor \frac{n}{r} \right\rfloor\right\}, T_{r,i} = \sum_{k=(i-1)r+1}^{ir} I_k.$$

Again, let us choose sequences $(\widehat{T}_{r,i})_{1\leq i\leq \lfloor n/r\rfloor}$ and $(\widehat{T}^{(m)}_{r,i})_{1\leq i\leq \lfloor n/r\rfloor}$ such that $\left((\widehat{T}_{r,i}, \widehat{T}^{(m)}_{r,i})\right)_{1\leq i\leq \lfloor n/r\rfloor}$ are i.i.d. replicas of $(T_{r,1}, T^{(m)}_{r,1})$. As in Barbour *et al.* [7], the compound Poisson parameters are $\theta_r = \frac{q_r}{rp}$, where $q_r = \mathbb{P}(T_{r,1} \geq 1)$, and $\nu_r = \mathcal{L}\left(T_{r,1} | \, T_{r,1} \geq 1\right)$, and we shall calculate them for this new configuration in the next section. The triangular inequality used is the same as in the last chapter;

$$
\begin{aligned}
& d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\
\leq\ & d_W\left(\mathcal{L}\big(N_n(\bullet)\big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T_{r,i}\delta_{irp}(\bullet)\Big)\right) && (6.1)\\
+\ & d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T^{(m)}_{r,i}\delta_{irp}(\bullet)\Big)\right) && (6.2)\\
+\ & d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T^{(m)}_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\Big)\right) && (6.3)\\
+\ & d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\Big)\right) && (6.4)\\
+\ & d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\Big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\right). && (6.5)
\end{aligned}
$$

However this time we shall need new arguments to bound the distances (6.2) to (6.4).

6.1 Link between a d-th order Markov chain and its related compound Poisson measure

We are now ready to evaluate the summands of the triangular inequality. The first result is the analogue of Lemma 5.2.1 for d-th order Markov chains, and provides a bound on (6.2).

Lemma 6.1.1 *With the above notations, we have*

$$
\begin{aligned}
& d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T_{r,i}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r \rfloor} T^{(m)}_{r,i}\delta_{irp}(\bullet)\Big)\right) \\
\leq\ & 1-(1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(m-d)\lfloor \frac{n}{r} \rfloor}\left(\prod_{j=1}^{d}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor \frac{n}{r} \rfloor - 1}.
\end{aligned}
$$

Proof. As in the proof of Lemma 5.2.1, we have

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\right)\right) \leq 1-\mathbb{P}\left(\forall\, k\in\overline{B}, I_k=0\right)$$

where $\overline{B}=\bigcup_{i=1}^{\lfloor n/r\rfloor}\{ir-m+1,\ldots,ir\}$.

Now, $\left\{\forall\, k\in\overline{B}, I_k=0\right\}$ is the situation in which all the elements of the d-th order Markov chain $\left(I_k\right)_{1\leq k\leq n}$ are equal to 0 on the last m spots of each block. To obtain such a configuration, we first need to have 0 on the "first" element of $\overline{B}$, which happens with probability $1-p$, then 0 on the next one, which adds a multiplicative factor of $(1-p_{\boldsymbol{v}_1})=\mathbb{P}(I_2=0|I_1=0)$ to the probability, then 0 on the next one, which adds another multiplicative factor of $(1-p_{\boldsymbol{v}_2})=\mathbb{P}(I_3=0|I_2=0,I_1=0)$ to the probability, and so on until the $(r-m+d)$-th element of the first block. Then, we need the last $m-d$ elements of the first block to be put to zero. This adds a factor $(1-p_{\boldsymbol{v}_d})^{(m-d)}=\left(\mathbb{P}(I_{d+1}=0|I_d=0,\ldots,I_1=0)\right)^{(m-d)}$ to the probability. Then, we need to have the first of the m last elements in the second block equal to zero. This jump of $r-m+1$ lags adds a multiplicative factor of $(1-p_{\boldsymbol{w}_1})=\mathbb{P}(I_t=0|I_{t-r+m}=0,I_{t-r+m-1}=0,\ldots,I_{t-r+m-d+1}=0)$ to the probability. Then, the second of the last m elements in the second block needs to be put to zero. That is, given that the previous event is zero, and that $d-2$ consecutive events, the last of which is happening just $r-m+2$ lags before, are zeros as well. This adds a factor $(1-p_{\boldsymbol{w}_2})=\mathbb{P}(I_t=0|I_{t-1}=0,I_{t-r+m-1}=0,I_{t-r+m-2}=0,\ldots,I_{t-r+m-d+2}=0)$ to the probability. Once the whole "tail" of zeros has jumped over the $r-m+1$ first elements of the second block, a multiplicative factor $\prod_{j=1}^{d}(1-p_{\boldsymbol{w}_j})$ has been added to the probability. Carrying on like this, and given that there are $\lfloor\frac{n}{r}\rfloor$ blocks which imply $\lfloor\frac{n}{r}\rfloor-1$ jumps of $r-m+1$ lags, we end up with

$$\mathbb{P}\left(\forall\, k\in\overline{B}, I_k=0\right)=(1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(m-d)\lfloor\frac{n}{r}\rfloor}\left(\prod_{j=1}^{d}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor\frac{n}{r}\rfloor-1}.$$

Hence the result. □

The second result is the analogous for d-th order Markov chains of Lemma 5.2.2.

Lemma 6.1.2 *With the same notation, we have*

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}_{r,i}\delta_{irp}(\bullet)\right),\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right)$$
$$\leq\ 1-\left((1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(m-d)}\right)^{\lfloor\frac{n}{r}\rfloor}.$$

Proof. As in the proof of Lemma 5.2.2, we have

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}_{r,i}\delta_{irp}(\bullet)\right),\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right)$$
$$\leq\ 1-\left(\mathbb{P}\left(\forall\, k\in[\![r-m+1,r]\!],I_k=0\right)\right)^{\lfloor n/r\rfloor}.$$

We are again looking at the situation when the m last elements of the first block are equal to 0. To obtain such a configuration, we first need to have 0 on the "first" element of $\overline{B}$, which happens with probability $1-p$, then 0 on the next one, which adds a multiplicative factor of $(1-p_{\boldsymbol{v}_1}=\mathbb{P}(I_2=0|I_1=0))$ to the probability, then 0 on the next one, which adds another multiplicative factor of $(1-p_{\boldsymbol{v}_2})=\mathbb{P}(I_3=0|I_2=0,I_1=0)$ to the probability, and so on until the $(r-m+d)$-th element of the first block. Then, we need the last $m-d$ elements of the first block to be put to zero. This adds a factor $(1-p_{\boldsymbol{v}_d})^{(m-d)}=\left(\mathbb{P}(I_{d+1}=0|I_d=0,\ldots,I_1=0)\right)^{(m-d)}$ to the probability. Consequently, we end up with

$$\mathbb{P}\left(\forall\, k\in[\![r-m+1,r]\!],I_k=0\right)\ =\ (1-p)(1-p_{\boldsymbol{v}_d})^{(m-d)}\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i}).$$

Hence the result. □

The following result is the equivalent for d-th order Markov chains of Lemma 5.2.3.

Lemma 6.1.3 *Still with the same notations, and with all the assumptions on our underlying perpetuity, there exist constants $C_d\in\mathbb{R}_+$ and $\rho_d\in]0,1[$ such that*

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}T^{(m)}_{r,i}\delta_{irp}(\bullet)\right),\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\widehat{T}^{(m)}_{r,i}\delta_{irp}(\bullet)\right)\right)\ \leq\ \frac{n}{r}C_d\rho_d^m\,.$$

Proof. We know from [7] that, using Lindeberg's method of decomposition [48], and Berbee's Lemma [14],

$$d_W\left(\mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m)}\delta_{irp}(\bullet)\Big), \mathcal{L}\Big(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m)}\delta_{irp}(\bullet)\Big)\right) \leq \frac{n}{r}\beta_I(m),$$

where β_I is the mixing coefficient of absolute regularity of the d-th order Markov chain,

$$\beta_I(m) = \max_{1\leq k\leq n} \mathbb{E} \sup_{H\in\mathcal{F}_{k+m+1,n}^d} \big|\,\mathbb{P}\,(H|\mathcal{F}_{1,k}^d) - \mathbb{P}\,(H)\,\big|,$$

where $\forall\,(s,t)\in[\![1,n]\!]^2$, $s\leq t$, $\mathcal{F}_{s,t}^d = \sigma\big(\{I_k,\, s\leq k\leq t\}\big)$. Now, (I_t) is a d-th order Markov chain, hence if with denote $\boldsymbol{I}_t = (I_t,\ldots,I_{t+d-1})$, $(\boldsymbol{I}_t)_{t\in\mathbb{N}^*}$ forms a first-order Markov chain on the state space $\{0,1\}^d$. The assumption, in the beginning of this chapter, that all d-th order transition probabilities of (I_t) are strictly positive implies that any subset of the state space $\{0,1\}^d$ can be reached by $(\boldsymbol{I}_t)$ in at least d steps from anywhere, which implies that the chain is ψ-irreducible, and that $\{0,1\}^d$ is small, and hence petite, for $(\boldsymbol{I}_t)_{t\in\mathbb{N}^*}$. Let us now investigate the period of $(\boldsymbol{I}_t)$. We are in a situation when the state space is countable, so we can recall from Section 2.4 that if $\boldsymbol{i}\in\{0,1\}^d$, the period of $\boldsymbol{i}$ is the greater common divisor of $\{n\geq 1 : \mathbb{P}(\boldsymbol{I}_n = \boldsymbol{i}|\boldsymbol{I}_0 = \boldsymbol{i}) > 0\}$. Let $\boldsymbol{i}\in\{0,1\}^d$. By construction, $(\boldsymbol{I}_t)$ can reach any point of the state space in d steps, and also in $d+1$ steps. Hence, starting from $\boldsymbol{i}$, it can return there in d steps, and also in $d+1$ steps. This means that d and $d+1$ are in $\{n\geq 1 : \mathbb{P}(\boldsymbol{I}_n = \boldsymbol{i}|\boldsymbol{I}_0 = \boldsymbol{i}) > 0\}$, an therefore its greater common divisor is 1. This is true for all $\boldsymbol{i}$ in $\{0,1\}^d$, hence $(\boldsymbol{I}_t)$ is aperiodic. Consequently, Theorem 16.2.2 in [51] can be applied, and it shows that $(\boldsymbol{I}_t)_{t\in\mathbb{N}^*}$ is uniformly ergodic, and hence geometrically ergodic. Section 2.4.2 shows then that it is also geometrically absolutely regular. In other words, there exist $R_d > 0$ and $\rho_d\in]0,1[$ such that $\beta_{\boldsymbol{I}}(m)\leq R_d\rho_d^m$, where $\beta_{\boldsymbol{I}}(m)$ is the β-mixing coefficient of $(\boldsymbol{I}_t)_{t\in\mathbb{N}^*}$. Now, if $s<t$, $\sigma(\boldsymbol{I}_s,\ldots,\boldsymbol{I}_t) = \sigma\big((I_s,\ldots,I_{s+d-1}),\ldots,(I_t,\ldots,I_{t+d-1})\big)$. As a result, $\sigma(I_s,\ldots,I_t)\subset\sigma(\boldsymbol{I}_s,\ldots,\boldsymbol{I}_{t-d+1})$. At this stage, Lemma 2.4.1 tells us that $\beta_I(m)\leq\beta_{\boldsymbol{I}}(m+d)$. Putting $C_d = R_d\rho_d^d$, we get $\beta_I m\leq C_d\rho_d^m$. Hence the result. □

Now, as in the case of the first-order Markov chain, we can now regroup the results into an overall approximation bound, noting that $2rp$ and $q_r(1.65/\sqrt{1-q_r} + e^{q_r})$ still bound (6.1), respectively (6.5) for the same reasons as they bounded (5.1), respectively (5.5), where the explicit expression of $q_r = \mathbb{P}(T_{r,1} \geq 1)$ is given below.

Theorem 6.1.1 *Still with the same notation,*

$$\begin{aligned} & d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\ \leq\ & 2rp + \frac{nC_d}{r}\rho_d^m \\ & +1-(1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(m-d)\lfloor\frac{n}{r}\rfloor}\left(\prod_{j=1}^{d}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor\frac{n}{r}\rfloor-1} \\ & +1-\left((1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(m-d)}\right)^{\lfloor\frac{n}{r}\rfloor} + q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right), \end{aligned}$$

where

$$\begin{aligned} q_r &= 1-(1-p)(1-p_{\boldsymbol{v}_d})^{(r-d)}\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i}) \\ &= 1-(1-p)\Big(\mathbb{P}\big(I_{d+1}=0\big|I_d=0,\ldots,I_1=0\big)\Big)^{(r-d)} \\ &\quad \times\prod_{i=1}^{d-1}\mathbb{P}\big(I_{d+1}=0\big|I_d=0,\ldots,I_{d-i+1}=0\big). \end{aligned}$$

6.2 Empirical compound Poisson parameters

The relation $\theta_r = \frac{q_r}{rp}$ and the expression of q_r from Theorem 6.1.1 immediately give us the value of the Poisson intensity.

$$\boxed{\theta_r = \frac{1}{rp}\Big(1-(1-p)\prod_{i=1}^{d-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_d})^{(r-d)}\Big).}$$

Let us now focus on the empirical compounding distribution ν_r. As in the section concerning first-order Markov chains, if T is a random variable with distribution ν_r,

and if $k \in \mathbb{N}$, our aim is to find $\mathbb{P}(T = k)$. We have $\mathbb{P}(T = k) = 0$ for $k = 0$ and $k \geq r + 1$. Now, $\nu_r = \mathcal{L}(T_{r,1} | T_{r,1} \geq 1)$, hence for $k \in [\![1, r]\!]$, $\mathbb{P}(T = k) = \frac{1}{q_r}\mathbb{P}\left(\sum_{s=1}^{r} I_s = k\right)$, where (I_s) is a d-th order Markov chain with state space $\{0, 1\}$. The situation is now the following. In a sequence of size r from a d-th order Markov chain, with two possible states 0 and 1, if $k \in [\![1, r]\!]$, what is the probability of getting the output 1 exactly k times? The answer to this can also be derived from Goodman [38]. The result we shall now use concerns the distribution of the frequency matrix of $(d+1)$-tuples of consecutive outcomes. It is derived from the original theorem concerning first-order Markov chain, after having noted that if $(Y_t)_{t \in \mathbb{N}^*}$ is a d-th order Markov chain on $\{0, 1\}$, then $\left((Y_t, \ldots, Y_{t+d-1})\right)_{t \in \mathbb{N}^*}$ forms a first-order Markov chain on $\{0, 1\}^d$. The transition matrix of the latter is of size $2^d \times 2^d$, and given that in one transition, the last $d-1$ covariates of the starting point must equal the first $d-1$ covariates of the destination point, there are only two possible destinations for a given start, corresponding to the two possible value of the last digit. More formally, let us introduce the notations of the theorem. Let $Y_1, \ldots, Y_N$ be a finite sequence from a d-th order Markov chain with q different states. Then $(Y_1, \ldots, Y_d)$, ..., $(Y_{N-d+1}, \ldots, Y_N)$ is a finite sequence from a first-order Markov chain with q^d different states. Denoting $\boldsymbol{i} = (i_1, \ldots, i_d)$, and $\boldsymbol{j} = (j_1, \ldots, j_d)$, the one-step transition from $\boldsymbol{i}$ to $\boldsymbol{j}$ is only possible if $(i_2, \ldots, i_d) = (j_1, \ldots, j_{d-1})$. Only in such a case can the frequency of this transition, *i.e.* the frequency of the $(d+1)$-tuple $(i_1, \ldots, i_d, j_d)$ in the original chain, be non-zero, and it will be denoted $F_{\boldsymbol{i},\boldsymbol{j}}$ or $F_{\boldsymbol{i},j_d}$. Namely, we have

$$F_{\boldsymbol{i},j} = \sum_{s=1}^{N-d+1} \mathbb{I}_{\left\{(Y_s,\ldots,Y_{s+d})=(i_1,\ldots,i_d,j)\right\}}.$$

Let us now call $F = [F_{\boldsymbol{i},\boldsymbol{j}}]$ the $q^d \times q^d$ matrix of frequencies, $F_{\boldsymbol{i}\cdot} = \sum_{j=1}^{q} F_{\boldsymbol{i},j}$, and $P = [p_{\boldsymbol{i},\boldsymbol{j}}]$ the $q^d \times q^d$ transition matrix. F is a random $q^d \times q^d$ matrix, which coefficients take values in $[\![0, N-d+1]\!]$. The realisation of F at a given point ω of the probability space Ω is denoted by $F(\omega) = f$. For the sake of simplicity, we shall directly put $q = 2$, given that the state space is $\{0, 1\}$ in our case.

Theorem 6.2.1 (Goodman) *For $(\boldsymbol{w},\boldsymbol{v})$ in $\{0,1\}^d \times \{0,1\}^d$, let $T(F,\boldsymbol{w},\boldsymbol{v})$ be the $(\boldsymbol{w},\boldsymbol{v})$-th cofactor of the matrix $\widehat{M}=[\delta_{i,j}-F_{i,j}/F_{i\cdot}]_{(i,j)\in\{0,1\}^d\times\{0,1\}^d}$. Then*

$$\begin{aligned}\mathbb{P}\Big(\big(F,(Y_{N-d+1},\ldots,Y_N)\big)=(f,\boldsymbol{w})\Big|(Y_1,\ldots,Y_d)=\boldsymbol{v}\Big)\\ = T(f,\boldsymbol{w},\boldsymbol{v})\,\frac{\prod_{i\in\{0,1\}^d} f_{i\cdot}!}{\prod_{i\in\{0,1\}^d}\prod_{j=0}^{1} f_{i,j}!}\prod_{i\in\{0,1\}^d}\prod_{j=0}^{1} p_{i,j}^{f_{i,j}}.\end{aligned}$$

Given that the cofactor $T(F,\boldsymbol{w},\boldsymbol{v})$ is the same for the whole $\boldsymbol{w}$-th row (*c.f.* Goodman), it is independent of $\boldsymbol{v}$ and can be denoted by $T(F,\boldsymbol{w})$. In addition, we are dealing with a sequence $(Y_1,\ldots,Y_r)$ of length $N=r$. Hence we get

$$\begin{aligned}\mathbb{P}\Big(\big(F,(Y_{N-d+1},\ldots,Y_N)\big)=(f,\boldsymbol{w})\Big|(Y_1,\ldots,Y_d)=\boldsymbol{v}\Big)\\ = T(f,\boldsymbol{w})\,\frac{\prod_{i\in\{0,1\}^d}(f_{i,0}+f_{i,1})!}{\prod_{i\in\{0,1\}^d} f_{i,0}!f_{i,1}!}\prod_{i\in\{0,1\}^d} p_{i,0}^{f_{i,0}}p_{i,1}^{f_{i,1}}.\end{aligned}$$

Now, we are interested in the distribution of the number of occurences of 1's in the sequence $(Y_1,\ldots,Y_r)$. Let us denote this number by F_1. We have

$$F_1=\sum_{t=1}^{r}\mathbb{I}_{\{Y_t=1\}} = \sum_{i_1\ldots i_d\in\{0,1\}} F_{1,i_1,\ldots,i_d} + \sum_{t=r-d+1}^{r}\mathbb{I}_{\{Y_t=1\}}.$$

Indeed, F_1 can be also viewed as the sum of the number of $(d+1)$-tuples starting with a 1 and of the number of 1 in the d last elements of the sequence. Now, we want to know $\mathbb{P}(F_1=k)$, for $k\in[\![1,r]\!]$.

$$\begin{aligned}\mathbb{P}(F_1=k) &= \sum_{i_1\ldots i_d\in\{0,1\}}\mathbb{P}\Big(\{F_1=k\}\cap\big\{(Y_{r-d+1},\ldots,Y_r)=(i_1,\ldots,i_d)\big\}\Big)\\ &= \sum_{i_1\ldots i_d\in\{0,1\}}\mathbb{P}\Bigg(\Big\{\sum_{j_1\ldots j_d\in\{0,1\}}F_{1,j_1,\ldots,j_d}=k-\sum_{t=1}^{d}i_t\Big\}\\ &\qquad\cap\big\{(Y_{r-d+1},\ldots,Y_r)=(i_1,\ldots,i_d)\big\}\Bigg)\\ &= \sum_{f}\sum_{i_1\ldots i_d\in\{0,1\}}\mathbb{P}\Big(\sum_{j_1\ldots j_d\in\{0,1\}}F_{1,j_1,\ldots,j_d}=k-\sum_{t=1}^{d}i_t\Big|F=f\Big)\\ &\qquad\times\mathbb{P}\Big(\{F=f\}\cap\big\{(Y_{r-d+1},\ldots,Y_r)=(i_1,\ldots,i_d)\big\}\Big).\end{aligned}$$

Now, multiplying by $\mathbb{P}\left(\sum_{j_1\dots j_d\in\{0,1\}} F_{1,j_1,\dots,j_d} = k - \sum_{t=r-d+1}^{r} \mathbb{I}_{\{i_t=1\}} \middle| F = f\right)$ boils down to only considering the set $\mathbb{N}_d(i_1,\dots,i_d)$ of collections $\mathcal{C}(f)$ of 2^{d+1} integers denoted by $\mathcal{C}(f) = \left\{f_{j_1,\dots,j_{d+1}}, j_1\dots j_{d+1}\in\{0,1\}\right\}$ such that

- $\displaystyle\sum_{j_1\dots j_{d+1}\in\{0,1\}} f_{j_1,\dots,j_{d+1}} = r-d$
- $\displaystyle\sum_{j_2\dots j_{d+1}\in\{0,1\}} f_{1,j_2,\dots,j_{d+1}} = k - \sum_{s=1}^{d} i_s.$

For each collection of coefficients $\mathcal{C}(f)$, the related $2^d\times 2^d$ integer matrix of frequencies f is defined by the following. For all $\big((a_1,\dots,a_d),(b_1,\dots,b_d)\big)$ in $\{0,1\}^d\times\{0,1\}^d$, the $\big((a_1,\dots,a_d),(b_1,\dots,b_d)\big)$-th coefficient of f is $\delta_{(a_2,\dots,a_d),(b_1,\dots,b_{d-1})} f_{a_1,\dots,a_d,b_d}$. Hence we get

$$\mathbb{P}(F_1=k) = \sum_{i_1\dots i_d\in\{0,1\}} \sum_{\mathcal{C}(f)\in\mathbb{N}_d(i_1,\dots,i_d)} T\big(f,(i_1,\dots,i_d)\big)\, \frac{\prod_{\boldsymbol{i}\in\{0,1\}^d}(f_{\boldsymbol{i},0}+f_{\boldsymbol{i},1})!}{\prod_{\boldsymbol{i}\in\{0,1\}^d} f_{\boldsymbol{i},0}!f_{\boldsymbol{i},1}!} \prod_{\boldsymbol{i}\in\{0,1\}^d} p_{\boldsymbol{i},0}^{f_{\boldsymbol{i},0}} p_{\boldsymbol{i},1}^{f_{\boldsymbol{i},1}}.$$

Remark. No easy simplification of this formula seems to be available.

Chapter 7

One-dimensional perpetuities

Let us now specify which sort of time series this study is aimed at. Theories sketched in the previous sections will be applied to random sequences from the insurance and banking world. The application of extreme value theory to financial processes, such as GARCH or ARMA, can be found in Embrechts *et al.* [30]. These models are defined by a random recurrence equation, an evolution equation for the conditional mean in the case of ARMA processes, and for the conditional variance in the case of GARCH processes.

7.1 Perpetuities

Definition 7.1.1 (Stochastic recursion) *A stochastic recursion is a random process generating variables* $(X_t)_{t\in\mathbb{N}^*}$ *as follows, given an initial variable* X_0

$$\forall\, t \in \mathbb{N}^*, \quad X_t = B_t X_{t-1} + A_t \tag{7.1}$$

where $((A_t, B_t))_{t\in\mathbb{N}^*}$ *is an i.i.d. sequence independent of* X_0.

Perpetuities are the resulting random sums, of the form

$$X_t = \sum_{m=1}^{t} A_m \prod_{j=m+1}^{t} B_j + X_1 \prod_{j=1}^{t} B_j$$

with the convention $\prod_{i=a}^{b} X_i = 1$ whenever $b < a$. See, for instance, the overview paper by Embrechts and Goldie [29], as well as Dufresne [28] for an application to risk theory and pension funding. The concept of "perpetuity", comes from the term "perpetual payment streams" in the insurance world. If an agent invests some money A_t at the beginning of each period, and the accumulated payments X_{t-1} are subject to the interest rate B_t, then the accumulated payments X_t at the end of the current period have the form of (7.1).

Beyond this interpretation, these equations encompass a vast panel of situations, and are especially suited to describe financial behaviours. For instance, they are satisfied by AR(1) and ARCH(1) processes, and by the squared volatility of GARCH(1,1) processes, *c.f.* Section 4. Higher orders of such models can be dealt with by considering random vectors and matrices for the generating coefficients. Yet, MA(1) sequences cannot be viewed as perpetuities, even if the current work is nevertheless relevant for some of them. First-order stochastic recursions are actually generalisations of AR(1) models where the recurrence coefficients are random.

7.2 Stationarity

Once this framework is set up, a natural question that arises is the existence of a stationary solution to such stochastic recurrence equations. In other words, can we find a marginal distribution for X_0 that would make $(X_t)_{t\in\mathbb{N}^*}$ a strictly stationary sequence, and what would be the shape of such a solution? This has been answered by Kesten [44] in 1973, who had noticed that light-tailed innovations $(B_t)_{t\in\mathbb{N}^*}$ created

heavy-tailed stochastic recurrent processes. Kesten gave the stationarity conditions on the stochastic coefficients $((A_t, B_t))_{t\in\mathbb{N}^*}$, as well as the shape of the marginal tail.

Theorem 7.2.1 (Kesten) *Let $\big((A_t, B_t)\big)_{t\in\mathbb{N}^*}$ be an i.i.d. sequence of non-negative random variables in $\mathbb{R}_+^2$. Suppose in addition that A_t is non-zero almost surely. Then, with the notation from Definition 7.1.1, denoting $(A, B) = (A_1, B_1)$, if there exists a stricly positive real number κ such that:*

- $\forall\, t\in\mathbb{N}^*$, X_0 *is independent of* (A_t, B_t)
- $\mathbb{E}\, B^\kappa = 1$
- $0 < \mathbb{E}\,(B^\kappa \ln B) < +\infty$
- $-\infty \leq \mathbb{E}\, \ln B < 0$
- $\mathbb{E}\, A^\kappa < +\infty$
- $\mathbb{E}\, \ln^+ A < +\infty$

then

i) $\displaystyle\sum_{m=1}^{t} A_m \prod_{j=1}^{m-1} B_j$ *converges in distribution to* $\displaystyle\sum_{m=1}^{+\infty} A_m \prod_{j=1}^{m-1} B_j$ *as* $t \to +\infty$

ii) $\displaystyle X \stackrel{d}{=} \sum_{m=1}^{+\infty} A_m \prod_{j=1}^{m-1} B_j$ *is the only distribution such that*

- $X_t \xrightarrow{d} X$
- $X \stackrel{d}{=} A + BX$
- X *is independent of* (A, B)

iii) *If* $X_0 \stackrel{d}{=} X$, *then* $(X_t)_{t\in\mathbb{N}^*}$ *is strictly stationary*

iv) $\mathbb{P}(X>x) \underset{x\to+\infty}{\sim} cx^{-\kappa}$ *with* $c = \dfrac{\mathbb{E}\left[((A+BX)^+)^\kappa - ((BX)^+)^\kappa\right]}{\kappa\,\mathbb{E}\,(|B|^\kappa \ln|B|)}$

This version of the theorem combines findings by Kesten and Goldie. Kesten found the existence of the constant κ in the general framework of random matrices, con-

firming his initial intuition, and Goldie [36], with renewal arguments, found the exact shape of the tail of a perpetuity, giving the explicit expressions of the constants κ as the solution to $\mathbb{E}\,B^{\kappa} = 1$ and c as in **iv)**. In particular, **iv)** shows that the resulting perpetuity is regularly varying. Vervaat [67] found properties of this limit distribution, including its exact expression when the coefficients are Γ- or β-distributed.

Remark. Now, a simpler version of this result, due to Brandt [19], can be formulated if one is only interested in the stationarity of the perpetuity, and not its tail, *c.f.*, for instance, Basrak *et al.* [12]. We need not assume the positiveness of the random coefficients to ensure stationarity. Indeed, if $\forall\, t \in \mathbb{N}^*$, X_0 is independent of (A_t, B_t), $\mathbb{E}\,\ln^+ A < +\infty$ and $-\infty \leq \mathbb{E}\,\ln B < 0$, then conclusions **i)**, **ii)** and **iii)** hold. These are the fundamental results, which allow us to apply extreme value theory to stochastic recursions.

7.3 Stability properties

It is quite easy to make our perpetuity Harris recurrent. For instance, this happens whenever $(A_t)_{t\in\mathbb{N}^*}$ and $(B_t)_{t\in\mathbb{N}^*}$ are independent of each other, and if both A and B have density functions which are strictly positive on the whole state space $S = [0, +\infty[$. In such a case, given that any region of S can therefore be reached from anywhere in one step, our perpetuity forms a ψ-irreducible, aperiodic Markov chain, for which the whole state space S is ν_1-small, and hence ν_{δ_1}-petite. Consequently, $(X_t)_{t\in\mathbb{N}^*}$ is uniformly ergodic. Putting $C = S$, which is petite, we have $\tau_C = 1$, and hence $\mathbb{P}(\tau_C < +\infty | X_0 = x) = 1$. Therefore, the resulting perpetuity is Harris recurrent.

The stochastic recursive nature of a perpetuity supplies it with interesting time-dependence features, which, again, we can assess with ergodicity and β-mixing. Basrak *et al.* [12] give conditions on the random coefficients for the resulting perpetuity

to be geometrically ergodic, using a result from Feigin and Tweedie [33] inspired by Nummelin and Tuominen [55]. They show that if $(X_t)_{t\in\mathbb{N}^*}$ verifies Equation (7.1) and the assumptions of Theorem 7.2.1, if (X_t) is ψ-irreducible, and if there is $\epsilon > 0$ such that $\mathbb{E}|B|^\epsilon < 1$ and $\mathbb{E}|A|^\epsilon < +\infty$, then (X_t) is geometrically ergodic. Given that under the aforementioned conditions, it is geometrically ergodic, using the result from Davydov [26] as in Section 2.4.2, we have the existence of two constants $\rho \in]0,1[$ and $R > 0$ such that $\beta(l) \leq R\rho^l$. It then follows from Lemma 2.4.1 that the mixing coefficient of the exceedance process has the same bound, *i.e.* $\beta_u(l) \leq R\rho^l$. Mokkadem [54] uses the same line of arguments presented in Nummelin and Tuominen [55], using petite sets and Harris recurrence, to show the geometric mixing of processes of the type $X_{t+1} = \varphi(X_t, e_t)$ on the state space $\mathbb{R}^d$, where (e_t) are i.i.d. and φ is a polynomial application. Veretennikov [66] gives a bound on the β-mixing coefficient of random processes verifying recurrence equations of the type $X_{t+1} = f(X_t) + W_t$ on the state space $\mathbb{R}^d$, where the (W_t) are i.i.d. and f is a mapping on $\mathbb{R}^d$ whose norm is bounded by classic functions specified in the paper. Yet, this does not apply to situations where f is a random function.

7.4 Distance to m-dependence of the perpetuity exceedances

Let us now start our approximation work. Because of the recursive essence of perpetuities, there is a way to relate their exceedances to an m-dependent process. Such processes are thereafter much easier to deal with. The following theorem is the first phase of the approximation.

Theorem 7.4.1 *Let $(X_t)_{1\leq t\leq n}$ be a random process on $\mathbb{R}$ verifying Equation (7.1). Assume that $\mathbb{E}\ \ln^+|A| < +\infty$, $-\infty \leq \mathbb{E}\ \ln|B| < 0$, and X_0 has the stationary distribution of Theorem 7.2.1, admits a bounded density function f_X, and has finite second*

moment $\mathbb{E}X^2$. *Then, for every* m *in* $[\![1,n]\!]$, *there exists an* m*-dependent process* $(I)_{1\leq t\leq n}$ *such that*

$$d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right),\mathcal{L}\left(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\right)\right)$$
$$\leq\ 8(n-m)\left\|f_X\right\|_\infty^{2/3}\left(\mathbb{E}X^2\right)^{1/3}\left(\mathbb{E}B^2\right)^{m/3},$$

where $\|\cdot\|_\infty$ *is the sup-norm.*

Proof. Let $m\in[\![1,n]\!]$. Because $(X_t)_{1\leq t\leq n}$ verifies Equation (7.1), we have, for all t in $[\![2,n]\!]$, for all m in $[\![1,t-1]\!]$,

$$X_t=B_tX_{t-1}+A_t\ \text{ and hence }\ X_t=\prod_{j=0}^{m-1}B_{t-j}X_{t-m}+\sum_{k=t-m+1}^{t}A_k\prod_{j=k+1}^{t}B_j. \quad (7.2)$$

Given that $\mathbb{E}\ \ln^+|A|<+\infty$ and $-\infty\leq\mathbb{E}\ \ln|B|<0$, the remark following Theorem 7.2.1 implies that

$$\sum_{k=1}^{+\infty}A_k\prod_{j=1}^{k-1}B_j$$

converges almost surely, and that if X_0 is chosen as a random variable independent of $\left((A_t,B_t)\right)_{1\leq t\leq n}$ with distribution law

$$\mathcal{L}(X_1)=\mathcal{L}\left(\sum_{k=1}^{+\infty}A_k\prod_{j=1}^{k-1}B_j\right),$$

then the resulting process $(X_t)_{1\leq t\leq n}$ is strictly stationary with the marginal distribution just specified. Now, let $\widehat{X}_1,\ldots,\widehat{X}_{n-m}$ be $n-m$ i.i.d. replicates of X_0. For t in $[\![1,m]\!]$, let $\widetilde{X}_t=X_t$. For t in $[\![m+1,n]\!]$, let

$$\widetilde{X}_t=\prod_{j=0}^{m-1}B_{t-j}\widehat{X}_{t-m}+\sum_{k=t-m+1}^{t}A_k\prod_{j=k+1}^{t}B_j.$$

In other words, this new process is built as if, for each lag, the recursion had started at the most m lags before, which is clear after comparison with Equation (7.2). The resulting process $(\widetilde{X}_t)_{1\leq t\leq n}$ is therefore m-dependent, since the (A_t,B_t) and the $(\widehat{X}_t)$

are i.i.d. Now, the remark following Theorem 7.2.1 implies that a sequence of length m of consecutive elements from $(\widetilde{X}_t)$ is strictly stationary, and given that the whole series is m-dependent, this ensures that the strict stationarity criterion is met. Indeed, a joint distribution of events which are spread over a period of time greater than m can then be split into the product of distributions of events which occur within m lags of each other. The resulting products are then invariant by translation of lags. Hence $(\widetilde{X}_t)$ is strictly stationary with marginal distribution law $\mathcal{L}(\widetilde{X}_t) = \mathcal{L}(X_t) = \mathcal{L}(X_1)$. Now, let $I_t = \mathbb{I}_{\{\widetilde{X}_t > u\}}$. It is clear that

$$\begin{aligned}
&\beta\big(\sigma(I_1,\ldots,I_t),\sigma(I_{t+m+1},\ldots,I_n)\big)\\
&\quad= \beta\Big(\sigma\big(\{\widetilde{X}_1>u\},\ldots,\{\widetilde{X}_t>u\}\big),\sigma\big(\{\widetilde{X}_{t+m+1}>u\},\ldots,\{\widetilde{X}_n>u\}\big)\Big).
\end{aligned}$$

We also know that

$$\sigma\big(\{\widetilde{X}_s>u\},\ldots,\{\widetilde{X}_t>u\}\big) \;\subset\; \sigma(\widetilde{X}_s,\ldots,\widetilde{X}_t),$$

and given Lemma 2.4.1, this implies

$$\begin{aligned}
&\beta\Big(\sigma\big(\{\widetilde{X}_1>u\},\ldots,\{\widetilde{X}_t>u\}\big),\sigma\big(\{\widetilde{X}_{t+m+1}>u\},\ldots,\{\widetilde{X}_n>u\}\big)\Big)\\
&\quad\le\; \beta\Big(\sigma(\widetilde{X}_1,\ldots,\widetilde{X}_t),\sigma(\widetilde{X}_{t+m+1},\ldots,\widetilde{X}_n)\Big).
\end{aligned}$$

Now, because $(\widetilde{X}_t)_{1\le t\le n}$ is m-dependent, the latter coefficient is zero, and hence $\beta\big(\sigma(I_1,\ldots,I_t),\sigma(I_{t+m+1},\ldots,I_n)\big) = 0$. Now, it follows from Remark 1 in Doukhan [27] that this implies the independence of $\sigma(I_1,\ldots,I_t)$ and $\sigma(I_{t+m+1},\ldots,I_n)$. Consequently, $(I_t)_{1\le t\le n}$ is also m-dependent, which is intuitively clear given its definition.

In addition, it is clear from Corollary 2.3.1 that

$$\begin{aligned}
& d_W\left(\mathcal{L}\Big(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\Big),\mathcal{L}\Big(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\Big)\right)\\
\leq\ & d_{TV}\left(\mathcal{L}\Big(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\Big),\mathcal{L}\Big(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\Big)\right)\\
\leq\ & \sum_{t=1}^{n}\mathbb{P}\left(\mathbb{I}_{\{X_t>u\}}\neq I_t\right)\\
\leq\ & \sum_{t=1}^{n}\mathbb{P}\left(\mathbb{I}_{\{X_t>u\}}\neq \mathbb{I}_{\{\widetilde{X}_t>u\}}\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(\mathbb{I}_{\{X_{m+1}>u\}}\neq \mathbb{I}_{\{\widetilde{X}_{m+1}>u\}}\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(X_{m+1}>u,\widetilde{X}_{m+1}\leq u\right)+(n-m)\,\mathbb{P}\left(X_{m+1}\leq u,\widetilde{X}_{m+1}>u\right),
\end{aligned}$$

as the first m elements are the same by construction. Now, let us denote

$$B_t^{(m)}=\prod_{j=0}^{m-1}B_{t-j}\quad\text{and}\quad A_t^{(m)}=\sum_{k=t-m+1}^{t}A_k\prod_{j=k+1}^{t}B_j,$$

so that

$$\widetilde{X}_t=B_t^{(m)}\widehat{X}_{t-m}+A_t^{(m)}\quad\text{and}\quad X_t=B_t^{(m)}X_{t-m}+A_t^{(m)}.$$

We then have

$$\begin{aligned}
& d_W\left(\mathcal{L}\Big(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\Big),\mathcal{L}\Big(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\Big)\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1>u-A_{m+1}^{(m)},B_{m+1}^{(m)}\widehat{X}_1\leq u-A_{m+1}^{(m)}\right)\\
& +(n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1\leq u-A_{m+1}^{(m)},B_{m+1}^{(m)}\widehat{X}_1>u-A_{m+1}^{(m)}\right).
\end{aligned}$$

Now, let $\epsilon_m>0$. Let us introduce a gap of length ϵ_m, and split our distribution function accordingly.

$$\begin{aligned}
& (n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1>u-A_{m+1}^{(m)},B_{m+1}^{(m)}\widehat{X}_1\leq u-A_{m+1}^{(m)}\right)\\
=\ & (n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1>u-A_{m+1}^{(m)}+\epsilon_m,B_{m+1}^{(m)}\widehat{X}_1\leq u-A_{m+1}^{(m)}\right)\\
& +(n-m)\,\mathbb{P}\left(u-A_{m+1}^{(m)}<B_{m+1}^{(m)}X_1\leq u-A_{m+1}^{(m)}+\epsilon_m,B_{m+1}^{(m)}\widehat{X}_1\leq u-A_{m+1}^{(m)}\right).
\end{aligned}$$

Hence, given that the marginal distribution f_X of the perpetuity is bounded,

$$\begin{aligned}
&(n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1 > u - A_{m+1}^{(m)}, B_{m+1}^{(m)}\widehat{X}_1 \le u - A_{m+1}^{(m)}\right)\\
&\le\ (n-m)\,\mathbb{P}\left(|B_{m+1}^{(m)}X_1 - B_{m+1}^{(m)}\widehat{X}_1| > \epsilon_m\right) + (n-m)\epsilon_m \left\|f_X\right\|_\infty\\
&\le\ (n-m)\,\mathbb{P}\left(|B_{m+1}^{(m)}|\cdot|X_1 - \widehat{X}_1| > \epsilon_m\right) + (n-m)\epsilon_m \left\|f_X\right\|_\infty.
\end{aligned}$$

Now, using Chebychev's inequality, we get

$$\begin{aligned}
&(n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1 > u - A_{m+1}^{(m)}, B_{m+1}^{(m)}\widehat{X}_1 \le u - A_{m+1}^{(m)}\right)\\
&\le\ (n-m)\,\frac{\mathrm{Var}\left(|B_{m+1}^{(m)}|\cdot|X_1 - \widehat{X}_1|\right)}{\epsilon_m^2} + (n-m)\epsilon_m \left\|f_X\right\|_\infty.
\end{aligned}$$

Given that, for A and B independent,

$$\mathrm{Var}AB \ \le\ \mathbb{E}(AB)^2 = \mathbb{E}A^2\mathbb{E}B^2,$$

we get

$$\mathrm{Var}\left(|B_{m+1}^{(m)}|\cdot|X_1 - \widehat{X}_1|\right) \ \le\ \mathbb{E}|B_{m+1}^{(m)}|^2\mathbb{E}|X_1 - \widehat{X}_1|^2.$$

Now, given that $B_{m+1}^{(m)} = \prod_{j=2}^{m+1} B_j$ with the B_j independent of each other, we have $\mathbb{E}|B_{m+1}^{(m)}|^2 = \left(\mathbb{E}B^2\right)^m$. In addition, $\mathbb{E}(X_1 - \widehat{X}_1)^2 \le 4\mathbb{E}X^2$. Finally, we obtain

$$\mathrm{Var}\left(|B_{m+1}^{(m)}|\cdot|X_1 - \widehat{X}_1|\right) \ \le\ 4\left(\mathbb{E}B^2\right)^m\mathbb{E}X^2.$$

We can now return to our approximation and substitute as follows.

$$\begin{aligned}
&(n-m)\,\mathbb{P}\left(B_{m+1}^{(m)}X_1 > u - A_{m+1}^{(m)}, B_{m+1}^{(m)}\widehat{X}_1 \le u - A_{m+1}^{(m)}\right)\\
&\le\ (n-m)\,\frac{4\left(\mathbb{E}B^2\right)^m\mathbb{E}X^2}{\epsilon_m^2} + (n-m)\epsilon_m \left\|f_X\right\|_\infty.
\end{aligned}$$

By symmetry between X and $\widehat{X}$, we can use an analogous argument to find the same bound on $n\,\mathbb{P}\left(B_{m+1}^{(m)}X_1 \le u - A_{m+1}^{(m)}, B_{m+1}^{(m)}\widehat{X}_1 > u - A_{m+1}^{(m)}\right)$. As a result,

$$\begin{aligned}
&d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), \mathcal{L}\left(\sum_{t=1}^{n} I_t\delta_{tp}(\bullet)\right)\right)\\
&\le\ (n-m)\,\frac{8\left(\mathbb{E}B^2\right)^m\mathbb{E}X^2}{\epsilon_m^2} + 2(n-m)\epsilon_m \left\|f_X\right\|_\infty.
\end{aligned}$$

This bound is meaningful, giving that all elements on the right-hand side are bounded, according to the assumptions of the theorem. Now, we can minimise it over the choices of ϵ_m. Denoting $g(\epsilon) = 8(n-m)\epsilon^{-2}\big(\mathbb{E}B^2\big)^m\mathbb{E}X^2 + 2(n-m)\epsilon\big\|f_X\big\|_\infty$, we get

$$\begin{aligned} g'(\epsilon^*) = 0 \quad &\Longleftrightarrow \quad -\frac{16(n-m)}{\epsilon^{*3}}\big(\mathbb{E}B^2\big)^m\mathbb{E}X^2 + 2(n-m)\big\|f_X\big\|_\infty = 0 \\ &\Longleftrightarrow \quad \epsilon^{*3} = \frac{8\big(\mathbb{E}B^2\big)^m\mathbb{E}X^2}{\big\|f_X\big\|_\infty} \\ &\Longleftrightarrow \quad \epsilon^* = 2\big(\mathbb{E}B^2\big)^{m/3}\big(\mathbb{E}X^2\big)^{1/3}\big\|f_X\big\|_\infty^{-1/3}. \end{aligned}$$

Now, $g''(\epsilon^*) = 48(n-m)(\epsilon^*)^{-4}\big(\mathbb{E}B^2\big)^m\mathbb{E}X^2 > 0$, hence ϵ^* minimizes the bound. Finally,

$$\begin{aligned} g(\epsilon^*) \quad = \quad & 8(n-m)\big(\mathbb{E}B^2\big)^m\mathbb{E}X^2 \cdot 2^{-2}\big(\mathbb{E}B^2\big)^{-2m/3}\big(\mathbb{E}X^2\big)^{-2/3}\big\|f_X\big\|_\infty^{2/3} \\ & +2(n-m)\big\|f_X\big\|_\infty \cdot 2\big(\mathbb{E}B^2\big)^{m/3}\big(\mathbb{E}X^2\big)^{1/3}\big\|f_X\big\|_\infty^{-1/3}. \end{aligned}$$

Adding the terms gives the result. □

Now, if we make the extra assumption that $\mathbb{E}B^2 < 1$, we can see that, choosing for instance $m = 3\sqrt{n}$, the bound tends to zero as n tends to infinity, which means this approximation makes sense!

Chapter 8

Multidimensional perpetuities

The multidimensional case is of particular importance in finance, simply because an investment portfolio of d assets forms a d-dimensional vector. The interest for investors is not only to look at the evolution of each asset separately, because this will not give them enough information to build the portfolio profit-and-loss distribution. Indeed it is only once they know how the components interact with each other that they are in a position to describe the portfolio properly. In this chapter, we do not look at individual marginal threshold exceedances any more. We consider the whole vector of assets, and its crossings over a d-dimensional border, given its stochastic recursive nature. The latter enables us to link it to a judicious m-dependent process. Together with the compound Poisson approximation that follows, this is intended as the main contribution of this work to the field of multivariate extreme value theory for stochastic recurrence equations.

8.1 Stochastic recursions

We are now looking at vectors of dimension d which verify a random recurrence equation of the first order, as follows:

$$\forall t \in \mathbb{N}^*, \quad \boldsymbol{Y}_t = \boldsymbol{B}_t \boldsymbol{Y}_{t-1} + \boldsymbol{A}_t, \tag{8.1}$$

where $\boldsymbol{Y}_t = (Y_{1,t}, \ldots, Y_{d,t})^{\mathrm{T}}$. Here $(\boldsymbol{B}_t)_{t\in\mathbb{N}^*}$ is a series of $d\times d$ random matrices, where $B_{i,j,t}$ is the coefficient of $\boldsymbol{B}_t$ at the intersection of row i and column j. $(\boldsymbol{A}_t)_{t\in\mathbb{N}^*}$ is a series of random vectors of dimension d, where $A_{i,t}$ is the i^{th} coefficient of $\boldsymbol{A}_t$. As a result, for all (i,j) in $[\![1,d]\!]^2$, $(Y_{i,t})_{t\in\mathbb{N}^*}$, $(B_{i,j,t})_{t\in\mathbb{N}^*}$ and $(A_{i,t})_{t\in\mathbb{N}^*}$ are real-valued stochastic processes. In the framework of stochastic recursions, we assume that $(\boldsymbol{A}_t, \boldsymbol{B}_t)_{t\in\mathbb{N}^*}$ are i.i.d. The state space $\mathbb{R}^d$ is equipped with the Euclidian norm $\|\cdot\|_2$, which naturally supplies the space $\mathrm{L}(\mathbb{R}^d)$ of endomorphisms of $\mathbb{R}^d$ with the operator norm $|||\cdot|||$, such that

$$\forall \boldsymbol{l} \in \mathrm{L}(\mathbb{R}^d), |||\boldsymbol{l}||| = \sup_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{l}(\boldsymbol{x})\|_2.$$

8.2 Stationarity

The stationary solution to (8.1) is given in Kesten [44], where it is shown that any linear combination of covariates from the resulting perpetuity is regularly varying in the univariate sense. Multivariate regular variations of perpetuities, where the dependence between covariates can be examined through the spectral measure of the whole vector, can be found in Stărică [64], *c.f.* also Pickands [57], de Haan [39] and Resnick [61] on the subject. Even though a more complete version is available in the literature, which is the analogue of Theorem 7.2.1 in a multivariate setting, we shall restrict ourselves to a modified version due to Brandt [19], as in Basrak *et al.* [12], which suffices for our present purposes, and which allows us not to assume the

positiveness of the random coefficients. Let $\left((\boldsymbol{A}_t, \boldsymbol{B}_t)\right)_{t\in\mathbb{N}*}$ be an i.i.d. sequence of d-dimensional random vectors $\boldsymbol{A}_t$ and $d\times d$ random matrices $\boldsymbol{B}_t$. First, if $\mathbb{E}\ln^+ |||\boldsymbol{B}_1||| < +\infty$, then $\frac{1}{n}\ln |||\boldsymbol{B}_1\cdots\boldsymbol{B}_n|||$ converges almost surely. Let us denote

$$\gamma = \lim_{n\to+\infty} \frac{1}{n}\ln |||\boldsymbol{B}_1\cdots\boldsymbol{B}_n|||;$$

γ is called the *Lyapunov exponent* of the sequence $(\boldsymbol{B}_t)_{t\in\mathbb{N}^*}$.

Theorem 8.2.1 (Brandt) *If* $\mathbb{E}\ln^+ |||\boldsymbol{B}_1||| < +\infty$, $\mathbb{E}\ln^+ \|\boldsymbol{A}_1\|_2 < +\infty$ *and* $\gamma < 0$, *then*

$$\sum_{k=1}^{+\infty}\prod_{j=1}^{k-1}\boldsymbol{B}_j\boldsymbol{A}_k$$

converges almost surely and is the unique strictly stationary solution of (8.1).

Remark. The condition $\gamma < 0$ holds if $\mathbb{E}\ln |||\boldsymbol{B}_1||| < 0$.

8.3 Markovian behaviour of the covariates

Let us now investigate how the behaviour of a covariate, say $(Y_{1,t})_{t\in\mathbb{N}^*}$, differs from the one described in the one-dimensional case, where the process of interest was simply $(Y_t)_{t\in\mathbb{N}^*}$. Let us then focus on the behaviour of $(Y_{1,t})_{t\in\mathbb{N}^*}$, and see what Equation (8.1) implies for this covariate. The following theorem shows that

Theorem 8.3.1 *If a d-dimensional vector verifies a stochastic recursion of order one, then each of its coefficients forms a d-th order Markov chain.*

Proof. Our stochastic recurrence equation is the following.

$$\begin{cases} Y_{1,t} = B_{1,1,t}Y_{1,t-1} + \cdots + B_{1,d,t}Y_{d,t-1} + A_{1,t} \\ Y_{2,t} = B_{2,1,t}Y_{1,t-1} + \cdots + B_{2,d,t}Y_{d,t-1} + A_{2,t} \\ \vdots \\ Y_{d,t} = B_{d,1,t}Y_{1,t-1} + \cdots + B_{d,d,t}Y_{d,t-1} + A_{d,t} \end{cases}$$

By continuity of the stochastic coefficients $B_{i,j,t}$ and $A_{i,t}$, this $d \times d$ linear system is non-singular almost surely, and we can "solve" the system of linear equations. In addition, the $B_{i,j,t}$ are equal to zero with probability 0. Hence, we can safely transform the last line as follows.

$$Y_{1,t} = \frac{1}{B_{d,1,t+1}} \left(Y_{d,t+1} - B_{d,2,t+1} Y_{2,t} - \cdots - B_{d,d,t+1} Y_{d,t} - A_{d,t+1} \right).$$

Now, let us substitute this expression of $Y_{1,t}$ into the $d-1$ first equations.

$$\begin{cases} \frac{1}{B_{d,1,t+1}} \left(Y_{d,t+1} - B_{d,2,t+1} Y_{2,t} - \cdots - B_{d,d,t+1} Y_{d,t} - A_{d,t+1} \right) = \\ \qquad \frac{B_{1,1,t}}{B_{d,1,t}} \left(Y_{d,t} - B_{d,2,t} Y_{2,t-1} - \cdots - B_{d,d,t} Y_{d,t-1} - A_{d,t} \right) \\ \qquad\qquad + B_{1,2,t} Y_{2,t-1} + \cdots + B_{1,d,t} Y_{d,t-1} + A_{1,t} & (1|1) \\ Y_{2,t} = \frac{B_{2,1,t}}{B_{d,1,t}} \left(Y_{d,t} - B_{d,2,t} Y_{2,t-1} - \cdots - B_{d,d,t} Y_{d,t-1} - A_{d,t} \right) \\ \qquad + B_{2,2,t} Y_{2,t-1} + \cdots + B_{2,d,t} Y_{d,t-1} + A_{2,t} & (2|1) \\ \vdots & \vdots \\ Y_{d-1,t} = \frac{B_{d-1,1,t}}{B_{d,1,t}} \left(Y_{d,t} - B_{d,2,t} Y_{2,t-1} - \cdots - B_{d,d,t} Y_{d,t-1} - A_{d,t} \right) \\ \qquad + B_{d-1,2,t} Y_{2,t-1} + \cdots + B_{d-1,d,t} Y_{d,t-1} + A_{d-1,t} & (d-1|1) \\ Y_{1,t} = \frac{1}{B_{d,1,t+1}} \left(Y_{d,t+1} - B_{d,2,t+1} Y_{2,t} - \cdots - B_{d,d,t+1} Y_{d,t} - A_{d,t+1} \right) & (1|*) \end{cases}$$

For $i \in [\![2, d-1]\!]$, equation $(i|1)$ can now be rewritten as

$$Y_{i,t} = \lambda_{i,2,t-1} Y_{2,t-1} + \cdots + \lambda_{i,d-1,t-1} Y_{d-1,t-1} + \lambda_{i,d,t-1} Y_{d,t-1} + \lambda_{i,d,t} Y_{d,t} + \lambda_{i,t}$$

where the $\lambda_{i,j,t}$ and $\lambda_{i,t}$ are random coefficients which are functions of only $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+1}$, $\boldsymbol{A}_t$ and $\boldsymbol{A}_{t+1}$. Now, substituting these expressions of $Y_{2,t}, \ldots, Y_{d-1,t}$ into $(1|1)$ proves the existence of random coefficients $\lambda_{1,j,t}$ and $\lambda_{1,t}$, which are functions of only $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+1}$, $\boldsymbol{B}_{t+2}$, $\boldsymbol{A}_t$, $\boldsymbol{A}_{t+1}$ and $\boldsymbol{A}_{t+2}$, such that

$$\begin{aligned} Y_{d-1,t} = {} & \lambda_{1,2,t} Y_{2,t} + \cdots + \lambda_{1,d-2,t} Y_{d-2,t} \\ & + \lambda_{1,d,t} Y_{d,t} + \lambda_{1,d,t+1} Y_{d,t+1} + \lambda_{1,d,t+2} Y_{d,t+2} + \lambda_{1,t}. \quad (d-1|*) \end{aligned}$$

Now, substituting this expression of $Y_{d-1,t}$ into equations $(2|1), \ldots, (d-2|1)$ enables us to find random coefficients $\eta_{i,j,t}$ and $\eta_{i,t}$, which are also functions of only $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+1}$,

$\boldsymbol{B}_{t+2}$, $\boldsymbol{A}_t$, $\boldsymbol{A}_{t+1}$ and $\boldsymbol{A}_{t+2}$, such that, for all $i \in [\![2, d-2]\!]$,

$$Y_{i,t} = \eta_{i,2,t-1}Y_{2,t-1} + \cdots + \eta_{i,d-2,t-1}Y_{d-2,t-1} + \eta_{i,d,t-1}Y_{d,t-1} + \eta_{i,d,t}Y_{d,t} + \eta_{i,d,t+1}Y_{d,t+1} + \eta_{i,t}. \quad (i|2)$$

Finally substituting this expression of $Y_{d-1,t}$ into equation $(d-1|1)$ gives us coefficients $\eta_{1,j,t}$ and $\eta_{1,t}$, which are functions of only $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+1}$, $\boldsymbol{B}_{t+2}$, $\boldsymbol{B}_{t+3}$, $\boldsymbol{A}_t$, $\boldsymbol{A}_{t+1}$, $\boldsymbol{A}_{t+2}$ and $\boldsymbol{A}_{t+3}$, such that

$$\lambda_{1,2,t}Y_{2,t} + \cdots + \lambda_{1,d-2,t}Y_{d-2,t} = \eta_{1,2,t-1}Y_{2,t-1} + \cdots + \eta_{1,d-2,t-1}Y_{d-2,t-1} + \eta_{1,d,t-1}Y_{d,t-1} + \eta_{1,d,t}Y_{d,t} + \eta_{1,d,t+1}Y_{d,t+1} + \eta_{1,d,t+2}Y_{d,t+2} + \eta_{1,t}. \quad (d-1|2)$$

Now, substituting the expressions of $Y_{2,t}, \ldots, Y_{d-2,t}$ from equations $(2|2), \ldots, (d-2|2)$ into equation $(d-1|2)$ gives us random coefficients $\epsilon_{d-2,j,t}$ and $\epsilon_{d-2,t}$, which are also functions of only $\boldsymbol{B}_t$, $\boldsymbol{B}_{t+1}$, $\boldsymbol{B}_{t+2}$, $\boldsymbol{B}_{t+3}$, $\boldsymbol{A}_t$, $\boldsymbol{A}_{t+1}$, $\boldsymbol{A}_{t+2}$ and $\boldsymbol{A}_{t+3}$, such that

$$Y_{d-2,t} = \epsilon_{d-2,2,t}Y_{2,t} + \cdots + \epsilon_{d-2,d-3,t}Y_{d-3,t} + \sum_{k=0}^{3} \epsilon_{d-2,d,t+k}Y_{d,t+k} + \epsilon_{d-2,t}. \quad (d-2|*)$$

Iterating this method, we can find a set of random coefficients $\epsilon_{i,j,t}$ and $\epsilon_{i,t}$, which are functions of only $\boldsymbol{B}_t$, ..., $\boldsymbol{B}_{t+d-i+1}$ and $\boldsymbol{A}_t$, ..., $\boldsymbol{A}_{t+d-i+1}$ such that, for for all $i \in [\![2, d-1]\!]$,

$$Y_{i,t} = \epsilon_{i,2,t}Y_{2,t} + \cdots + \epsilon_{i,i-1,t}Y_{i-1,t} + \sum_{k=0}^{d-i+1} \epsilon_{i,d,t+k}Y_{d,t+k} + \epsilon_{i,t}. \quad (i|*)$$

Then, substituting backwards in cascade the expressions of $Y_{i,t}$ into equations $(i+1|*)$, we obtain expressions of $Y_{2,t}, \ldots, Y_{d-1,t}$ as linear functions of the $d-1$ first lags of $Y_{d,t}$. Substituting one last time into equation $(1|1)$ proves that $Y_{d,t}$ verifies a d-th order recurrence equation of the following type:

$$Y_{d,t} = \alpha_{0,t} + \alpha_{1,t}Y_{d,t-1} + \cdots + \alpha_{d,t}Y_{d,t-d},$$

with random coefficients $\alpha_{i,t}$ which are functions of only $\boldsymbol{B}_{t-d}$, ..., $\boldsymbol{B}_t$ and $\boldsymbol{A}_{t-d}$, ..., $\boldsymbol{A}_t$. Now, given that $\big((\boldsymbol{A}_t, \boldsymbol{B}_t)\big)$ are i.i.d. by assumption, this implies that our random coefficients $(\alpha_{i,t})$ are at most d-dependent. Yet, for $i \neq j$, $\alpha_{i,t}$ and $\alpha_{j,t}$ might

be dependent. As a result, $\left(Y_{d,t}\right)_{t\in\mathbb{N}^*}$ forms a d-th order Markov chain. By symmetry, this is true for all the other covariates. □

This result now enables us to realise how the behaviour of the empirical exceedance point process of a one-dimensional random recursive variable, analysed in the previous chapters, is pertubated by the introduction of covariates.

8.4 Mixing behaviour of the exceedances

Let us now have a look at our multidimensional perpetuity from an extreme value theory viewpoint. In a first place we ought to define what we mean by threshold exceedance in a multivariate context. In the rest of this chapter, we shall consider vectors with non-negative entries, which means restraining our state space to $\mathbb{R}_+^d$. After having chosen thresholds $u_1,\ldots,u_d$ for our d covariates, for $\boldsymbol{u}=(u_1,\ldots,u_d)$, consider the d-dimensional parallelepiped

$$R_{0,\boldsymbol{u}} \;=\; [0,u_1]\times\cdots\times[0,u_d].$$

In addition, let us denote the layer $E_{0,\boldsymbol{u},\boldsymbol{v}} = R_{0,\boldsymbol{v}} \setminus R_{0,\boldsymbol{u}}$. We are looking at exceedances outside a d-dimensional parallelepiped, and if $(\boldsymbol{X}_t)_{1\leq t\leq n}$ is our perpetuity, the corresponding exceedance process is $\left(\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\right)_{1\leq t\leq n}$. Let us denote by $\beta_{\boldsymbol{u}}$ its absolute regularity coefficient. In order to describe its mixing behaviour, let us simply recall Theorem 2.8 in Basrak *et al.* [12], which we already used in the case of the one-dimensional perpetuity, but is stated in the multidimensional setting. Indeed, if $(\boldsymbol{X}_t)_{1\leq t\leq n}$ verifies Equation (8.1) and the assumptions of Theorem 8.2.1, if in addition $(\boldsymbol{X}_t)_{1\leq t\leq n}$ is ψ-irreducible and there exists $\epsilon>0$ such that $\mathbb{E}\|\boldsymbol{B}\|^{\epsilon}<1$ and $\mathbb{E}\|\boldsymbol{A}\|_2^{\epsilon}<+\infty$, then the resulting perpetuity is geometrically ergodic. Now, the result from Davydov [26] still holds, which means that the β-mixing coefficient $\beta_{\boldsymbol{X}}(m)$ of our multidimensional perpetuity can be rewritten as

$$\beta_{\boldsymbol{X}}(m)=\int_{\mathbb{R}_+^d} d_{TV}\Big(\mathbb{P}(\boldsymbol{X}_m\in\bullet|\boldsymbol{X}_0=\boldsymbol{x}),\pi(\bullet)\Big)\pi(\boldsymbol{dx}),$$

where π is the invariant distribution given by Theorem 8.2.1. Consequently, there exist $\rho \in]0,1[$ and $R > 0$ such that $\beta_{\boldsymbol{X}}(m) \leq R\rho^m$. Now, if $s < t$, we clearly have

$$\sigma\Big(\{\boldsymbol{X}_s \notin R_{0,\boldsymbol{u}}\}, \ldots, \{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}\Big) \subset \sigma(\boldsymbol{X}_s, \ldots, \boldsymbol{X}_t),$$

which implies, by Lemma 2.4.1, that

$$\beta\Big(\sigma\Big(\{\boldsymbol{X}_1 \notin R_{0,\boldsymbol{u}}\}, \ldots, \{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}\Big), \sigma\Big(\{\boldsymbol{X}_k \notin R_{0,\boldsymbol{u}}\}, k \geq t+m\Big)\Big) \leq \beta_{\boldsymbol{X}}(m),$$

and noting that

$$\begin{aligned}
&\beta\Big(\sigma\Big(\mathbb{I}_{\{\boldsymbol{X}_1 \notin R_{0,\boldsymbol{u}}\}}, \ldots, \mathbb{I}_{\{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}}\Big), \sigma\Big(\mathbb{I}_{\{\boldsymbol{X}_k \notin R_{0,\boldsymbol{u}}\}}, k \geq t+m\Big)\Big) \\
&\quad = \beta\Big(\sigma\Big(\{\boldsymbol{X}_1 \notin R_{0,\boldsymbol{u}}\}, \ldots, \{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}\Big), \sigma\Big(\{\boldsymbol{X}_k \notin R_{0,\boldsymbol{u}}\}, k \geq t+m\Big)\Big),
\end{aligned}$$

we finally obtain that $\beta_{\boldsymbol{u}} \leq \beta_{\boldsymbol{X}}(m)$, and hence $\beta_{\boldsymbol{u}} \leq R\rho^m$, which proves that the process of multidimensional threshold exceedances, under the assumptions above, is geometrically absolutely regular.

8.5 Link between multivariate exceedances and m-dependence

Let us now move on to the approximation of the point process of exceedances outside the multidimensional threshold $R_{0,\boldsymbol{u}}$. For this purpose, let us assume that the process $(\boldsymbol{X}_t)_{1 \leq t \leq n}$ follows Equation (8.1) and verifies the assumptions of Theorem 8.2.1 and $\mathbb{E} \ln \|\boldsymbol{B}_1\| < 0$, which ensure its strict stationarity. Now, for the following result to hold, we need to assume that $\boldsymbol{X}$, which has the stationary distribution of the perpetuity, as specified in Theorem 8.2.1, has non-negative coefficients. As a result, the state space considered here is not $\mathbb{R}^d$, but $\mathbb{R}^d_+$. This also implies that all random cofficients $(\boldsymbol{A}_t, \boldsymbol{B}_t)$ of the recursion must have non-negative entries.

Theorem 8.5.1 *Let us assume that* $\mathbb{E}\|\boldsymbol{X}\|_2^2 < +\infty$, *and that* $\boldsymbol{X}$ *has a bounded density function* $f_{\boldsymbol{X}}$ *on the state space* $S = \mathbb{R}_+^d$. *Then, for every* m *in* $[\![1, n]\!]$, *there exists an* m*-dependent process* $(I_t)_{1\le t\le n}$ *such that, for all* $\epsilon_m > 0$,

$$d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathcal{L}\left(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\right)\right)$$
$$\le \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_\infty,$$

where $\|\cdot\|_\infty$ *is the sup-norm.*

Proof. The beginning of the proof is similar to the one of Theorem 7.4.1. Let $m \in [\![1, n]\!]$. Because $(\boldsymbol{X}_t)_{1\le t\le n}$ verifies Equation (8.1), we have, for all t in $[\![2, n]\!]$, for all m in $[\![1, t-1]\!]$,

$$\boldsymbol{X}_t = \boldsymbol{B}_t\boldsymbol{X}_{t-1} + \boldsymbol{A}_t \implies \boldsymbol{X}_t = \prod_{j=0}^{m-1}\boldsymbol{B}_{t-j}\boldsymbol{X}_{t-m} + \sum_{k=t-m+1}^{t}\prod_{j=k+1}^{t}\boldsymbol{B}_j\boldsymbol{A}_k. \quad (8.2)$$

Given that $\mathbb{E}\,\ln^+\|\boldsymbol{A}_1\|_2 < +\infty$ and $\mathbb{E}\,\ln\|\boldsymbol{B}_1\| < 0$, Theorem 8.2.1 implies that

$$\sum_{k=1}^{+\infty}\prod_{j=1}^{k-1}\boldsymbol{B}_j\boldsymbol{A}_k$$

converges almost surely, and that if $\boldsymbol{X}_1$ is chosen as a random variable independent of $\left((\boldsymbol{A}_t, \boldsymbol{B}_t)\right)_{1\le t\le n}$ with distribution law

$$\mathcal{L}(\boldsymbol{X}_1) = \mathcal{L}\left(\sum_{k=1}^{+\infty}\prod_{j=1}^{k-1}\boldsymbol{B}_j\boldsymbol{A}_k\right),$$

then the resulting process $(\boldsymbol{X}_t)_{1\le t\le n}$ is strictly stationary with the marginal distribution just specified. Now, let $\widehat{\boldsymbol{X}}_1, \ldots, \widehat{\boldsymbol{X}}_{n-m}$ be $n-m$ i.i.d. replicates of $\boldsymbol{X}_1$. For t in $[\![1, m]\!]$, let $\widetilde{\boldsymbol{X}}_t = \boldsymbol{X}_t$. For t in $[\![m+1, n]\!]$, let

$$\widetilde{\boldsymbol{X}}_t = \prod_{j=0}^{m-1}\boldsymbol{B}_{t-j}\widehat{\boldsymbol{X}}_{t-m} + \sum_{k=t-m+1}^{t}\prod_{j=k+1}^{t}\boldsymbol{B}_j\boldsymbol{A}_k.$$

In other words, this new process is built as if, for each lag, the recursion had started at the most m lags before, which is clear after comparison with Equation (8.2).

The resulting process $(\widetilde{\boldsymbol{X}}_t)_{1\leq t\leq n}$ is therefore m-dependent, since the $(\boldsymbol{A}_t, \boldsymbol{B}_t)$ and the $(\widehat{\boldsymbol{X}}_t)$ are i.i.d. Now, Theorem 8.2.1 implies that a sequence of length m of consecutive elements from $(\widetilde{\boldsymbol{X}}_t)$ is strictly stationary, and given that the whole series is m-dependent, this ensures that the strict stationarity criterion is met. Indeed, a joint distribution of events which are spread over a period of time greater than m can then be split into the product of distributions of events which occur within m lags of eachother. The resulting products are then invariant by translation of lags. Hence $(\widetilde{\boldsymbol{X}}_t)$ is strictly stationary with marginal distribution law $\mathcal{L}(\widetilde{\boldsymbol{X}}_t) = \mathcal{L}(\boldsymbol{X}_t) = \mathcal{L}(\boldsymbol{X}_1)$. Now, given that we are in a multidimensional setting, our line of argument departs here slightly from the one in the proof of Theorem 7.4.1. Let $I_t = \mathbb{I}_{\{\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}}\}}$. It is clear that

$$\begin{aligned}&\beta\big(\sigma(I_1,\ldots,I_t),\sigma(I_{t+m+1},\ldots,I_n)\big)\\ &\quad= \beta\Big(\sigma\big(\{\widetilde{\boldsymbol{X}}_1 \notin R_{0,\boldsymbol{u}}\},\ldots,\{\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}}\}\big),\sigma\big(\{\widetilde{\boldsymbol{X}}_{t+m+1} \notin R_{0,\boldsymbol{u}}\},\ldots,\{\widetilde{\boldsymbol{X}}_n \notin R_{0,\boldsymbol{u}}\}\big)\Big).\end{aligned}$$

We also know that

$$\sigma\big(\{\widetilde{\boldsymbol{X}}_s \notin R_{0,\boldsymbol{u}}\},\ldots,\{\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}}\}\big) \;\subset\; \sigma(\widetilde{\boldsymbol{X}}_s,\ldots,\widetilde{\boldsymbol{X}}_t),$$

and given Lemma 2.4.1, this implies

$$\begin{aligned}&\beta\Big(\sigma\big(\{\widetilde{\boldsymbol{X}}_1 \notin R_{0,\boldsymbol{u}}\},\ldots,\{\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}}\}\big),\sigma\big(\{\widetilde{\boldsymbol{X}}_{t+m+1} \notin R_{0,\boldsymbol{u}}\},\ldots,\{\widetilde{\boldsymbol{X}}_n \notin R_{0,\boldsymbol{u}}\}\big)\Big)\\ &\qquad\leq\; \beta\Big(\sigma(\widetilde{\boldsymbol{X}}_1,\ldots,\widetilde{\boldsymbol{X}}_t),\sigma(\widetilde{\boldsymbol{X}}_{t+m+1},\ldots,\widetilde{\boldsymbol{X}}_n)\Big).\end{aligned}$$

Now, because $(\widetilde{\boldsymbol{X}}_t)_{1\leq t\leq n}$ is m-dependent, the latter coefficient is zero, and hence $\beta\big(\sigma(I_1,\ldots,I_t),\sigma(I_{t+m+1},\ldots,I_n)\big) = 0$. Now, it follows from Remark 1 in Doukhan [27] that this implies the independence of $\sigma(I_1,\ldots,I_t)$ and $\sigma(I_{t+m+1},\ldots,I_n)$. Consequently, $(I_t)_{1\leq t\leq n}$ is also m-dependent, which is intuitively clear from their definition.

In addition, it is clear from Corollary 2.3.1 that

$$\begin{aligned}
& d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathcal{L}\left(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\right)\right)\\
\leq\ & d_{TV}\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathcal{L}\left(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\right)\right)\\
\leq\ & \sum_{t=1}^{n}\mathbb{P}\left(\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\neq I_t\right)\\
\leq\ & \sum_{t=1}^{n}\mathbb{P}\left(\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\neq \mathbb{I}_{\{\widetilde{\boldsymbol{X}}_t\notin R_{0,\boldsymbol{u}}\}}\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(\mathbb{I}_{\{\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}}\}}\neq \mathbb{I}_{\{\widetilde{\boldsymbol{X}}_{m+1}\notin R_{0,\boldsymbol{u}}\}}\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)\\
& +(n-m)\,\mathbb{P}\left(\boldsymbol{X}_{m+1}\in R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\notin R_{0,\boldsymbol{u}}\right).
\end{aligned}$$

Now, let us denote

$$\boldsymbol{B}_t^{(m)}=\prod_{j=0}^{m-1}\boldsymbol{B}_{t-j}\quad\text{and}\quad\boldsymbol{A}_t^{(m)}=\sum_{k=t-m+1}^{t}\prod_{j=k+1}^{t}\boldsymbol{B}_j\boldsymbol{A}_k,$$

so that

$$\widetilde{\boldsymbol{X}}_t=\boldsymbol{B}_t^{(m)}\widehat{\boldsymbol{X}}_{t-m}+\boldsymbol{A}_t^{(m)}\quad\text{and}\quad\boldsymbol{X}_t=\boldsymbol{B}_t^{(m)}\boldsymbol{X}_{t-m}+\boldsymbol{A}_t^{(m)}.$$

Recall that we have

$$\begin{aligned}
& d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{0,\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathcal{L}\left(\sum_{t=1}^{n}I_t\delta_{tp}(\bullet)\right)\right)\\
\leq\ & (n-m)\,\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)\\
& +(n-m)\,\mathbb{P}\left(\boldsymbol{X}_{m+1}\in R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\notin R_{0,\boldsymbol{u}}\right).
\end{aligned}$$

Now, let $\epsilon_m>0$. Let us introduce a gap of length ϵ_m in all dimensions, and split our distribution function accordingly. Let $\boldsymbol{\epsilon}_m=(\epsilon_m,\ldots,\epsilon_m)\in\left(\mathbb{R}_+^*\right)^d$. Heuristically, if $\boldsymbol{X}_{m+1}$ and $\widetilde{\boldsymbol{X}}_{m+1}$ are within $\boldsymbol{\epsilon}_m$ of each other, then the probabilities above can be bounded using their common density. If they are farther than $\boldsymbol{\epsilon}_m$ from each other,

then a mixing argument will provide a bound.

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right) \qquad (8.3)\\ &= \mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}+\boldsymbol{\epsilon}_m},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)+\mathbb{P}\left(\boldsymbol{X}_{m+1}\in E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right).\end{aligned}$$

Hence, given that the marginal distribution $f_{\boldsymbol{X}}$ of the perpetuity is bounded, and that $E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}$ has a finite multidimensional volume equal to $\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i$, *c.f.* Remark 3 following the proof, we have

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)\\ &\leq \mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}+\boldsymbol{\epsilon}_m},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_\infty\\ &\leq \mathbb{P}\left(\left\|\boldsymbol{X}_{m+1}-\widetilde{\boldsymbol{X}}_{m+1}\right\|_2>\epsilon_m\right)+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_\infty\\ &\leq \mathbb{P}\left(\left\|\boldsymbol{B}_{m+1}^{(m)}\left(\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right)\right\|_2>\epsilon_m\right)+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_\infty.\end{aligned}$$

Owing to the essence of the operator norm on $\mathrm{L}(\mathbb{R}_+^d)$,

$$\left\|\boldsymbol{B}_{m+1}^{(m)}\left(\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right)\right\|_2 \leq |||\boldsymbol{B}_{m+1}^{(m)}|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2.$$

Another property of this norm is that, for $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ in $\mathrm{L}(\mathbb{R}_+^d)$,

$$|||\boldsymbol{l}_1\circ\boldsymbol{l}_2||| \leq |||\boldsymbol{l}_1|||\cdot|||\boldsymbol{l}_2|||,$$

and noting that $\boldsymbol{B}_{m+1}^{(m)}=\prod_{i=0}^{m-1}\boldsymbol{B}_{m+1-i}=\prod_{i=2}^{m+1}\boldsymbol{B}_i$, this implies

$$\left\|\boldsymbol{B}_{m+1}^{(m)}\left(\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right)\right\|_2 \leq \prod_{i=2}^{m+1}|||\boldsymbol{B}_i|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2.$$

As a result,

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)\\ &\leq \mathbb{P}\left(\prod_{i=2}^{m+1}|||\boldsymbol{B}_i|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2>\epsilon_m\right)+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_\infty,\end{aligned}$$

and using Chebychev's inequality, we get

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{0,\boldsymbol{u}},\,\widetilde{\boldsymbol{X}}_{m+1}\in R_{0,\boldsymbol{u}}\right)\\&\quad\leq\ \frac{1}{\epsilon_m^2}\operatorname{Var}\left(\prod_{i=2}^{m+1}|||\boldsymbol{B}_i|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2\right)+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\left\|f_{\boldsymbol{X}}\right\|_\infty.\end{aligned}$$

Noting that, for U and V independent, $\operatorname{Var}UV\leq\mathbb{E}(UV)^2=\mathbb{E}U^2\mathbb{E}V^2$, and that $|||\boldsymbol{B}_t|||,\ldots,|||\boldsymbol{B}_{t-m+1}|||,\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2$ are independent of eachother, we get

$$\operatorname{Var}\left(\prod_{i=0}^{m-1}|||\boldsymbol{B}_{t-i}|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2\right)\ \leq\ \left(\mathbb{E}|||\boldsymbol{B}|||^2\right)^m\mathbb{E}\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2^2.\tag{8.4}$$

Now, given that $\boldsymbol{X}_1$ and $\widehat{\boldsymbol{X}}_1$ are independent,

$$\begin{aligned}\mathbb{E}\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2^2\ &\leq\ \mathbb{E}\|\boldsymbol{X}_1\|_2^2+\mathbb{E}\|\widehat{\boldsymbol{X}}_1\|_2^2+2\mathbb{E}\|\boldsymbol{X}_1\|_2\mathbb{E}\|\widehat{\boldsymbol{X}}_1\|_2\\&\leq\ 2\mathbb{E}\|\boldsymbol{X}\|_2^2+2\Big(\mathbb{E}\|\boldsymbol{X}\|_2\Big)^2,\end{aligned}$$

and noting that $\mathbb{E}U^2\geq\left(\mathbb{E}U\right)^2$, we have

$$\mathbb{E}\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2^2\ \leq\ 4\mathbb{E}\|\boldsymbol{X}\|_2^2,$$

which we can substitute into (8.4) to obtain

$$\operatorname{Var}\left(\prod_{i=0}^{m-1}|||\boldsymbol{B}_{t-i}|||\cdot\left\|\boldsymbol{X}_1-\widehat{\boldsymbol{X}}_1\right\|_2\right)\ \leq\ 4\Big(\mathbb{E}|||\boldsymbol{B}|||^2\Big)^m\mathbb{E}\|\boldsymbol{X}\|_2^2.$$

Consequently,

$$\begin{aligned}&\mathbb{P}\left(\boldsymbol{X}_{m+1}\notin R_{\boldsymbol{0},\boldsymbol{u}},\,\widetilde{\boldsymbol{X}}_{m+1}\in R_{\boldsymbol{0},\boldsymbol{u}}\right)\\&\quad\leq\ \frac{4}{\epsilon_m^2}\Big(\mathbb{E}|||\boldsymbol{B}|||^2\Big)^m\mathbb{E}\|\boldsymbol{X}\|_2^2+\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\left\|f_{\boldsymbol{X}}\right\|_\infty.\end{aligned}$$

This bound is meaningful, giving that all elements on the right-hand side are bounded, according to the assumptions of the theorem. By symmetry between $\boldsymbol{X}_1$ and $\widetilde{\boldsymbol{X}}_1$, an analogous argument provides the same bound for $\mathbb{P}\left(\boldsymbol{X}_{m+1}\in R_{0,\boldsymbol{u}},\,\widetilde{\boldsymbol{X}}_{m+1}\notin R_{0,\boldsymbol{u}}\right)$ and completes the proof. □

Remark 1. Now, if we assume that $\mathbb{E}\|\|\boldsymbol{B}\|\|^2 < 1$, the first part of the bound, *i.e.* $\frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\|\boldsymbol{B}\|\|^2\right)^m \mathbb{E}\|\boldsymbol{X}\|_2^2$, decays geometrically, and choosing $\epsilon_m = 1/n^2$ and $m = \sqrt{n}$, the whole bound $8n^4(n-\sqrt{n})\left(\mathbb{E}\|\|\boldsymbol{B}\|\|^2\right)^{\sqrt{n}}\mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-\sqrt{n})\Big(\prod_{i=1}^d (u_i + 1/n^2) - \prod_{i=1}^d u_i\Big)\left\|f_{\boldsymbol{X}}\right\|_\infty$ becomes small for large values of n.

Remark 2. In the case when $d = 2$, which we shall illustrate numerically in Chapter 10, and considering exceedances outside $R_{0,\boldsymbol{u}} = [0, u_X] \times [0, u_Y]$ we obtain an m-dependent process $(I_t)_{1\leq t\leq n}$ such that

$$\begin{aligned} & d_W\left(\mathcal{L}\Big(\sum_{t=1}^n \mathbb{I}_{\{\boldsymbol{X}_t \notin R_{\boldsymbol{0},\boldsymbol{u}}\}}\delta_{tp}(\bullet)\Big), \mathcal{L}\Big(\sum_{t=1}^n I_t\delta_{tp}(\bullet)\Big)\right) \\ & \leq \frac{8(n-m)}{\epsilon_m^2}\Big(\mathbb{E}\|\|\boldsymbol{B}\|\|^2\Big)^m \mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-m)\epsilon_m(\epsilon_m + u_X + u_Y)\left\|f_{\boldsymbol{X}}\right\|_\infty. \end{aligned}$$

We can find an optimal value for ϵ_m. Denoting by $a = 8(n-m)\Big(\mathbb{E}\|\|\boldsymbol{B}\|\|^2\Big)^m \mathbb{E}\|\boldsymbol{X}\|_2^2$, $b = 2(n-m)\left\|f_{\boldsymbol{X}}\right\|_\infty$ and $c = u_X + u_Y$, we want to find ϵ which minimizes $g(\epsilon) = \frac{a}{\epsilon^2} + b\epsilon(\epsilon + c)$.

$$\begin{aligned} g'(\epsilon) = 0 \quad &\Longleftrightarrow \quad -\frac{2a}{\epsilon^3} + b(2\epsilon + c) = 0 \\ &\Longleftrightarrow \quad 2b\epsilon^4 + bc\epsilon^3 - 2a = 0. \end{aligned}$$

Solutions to this equation can then be found numerically.

Remark 3. In Equation (8.3), we use the fact that we are only looking at vectors with non-negative entries, which allows us to partition the state space $\mathbb{R}_+^d$ into $R_{0,\boldsymbol{u}}$, $E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}$ and $\mathbb{R}_+^d \setminus R_{0,\boldsymbol{u}+\boldsymbol{\epsilon}_m}$. Consequently, we argue that

$$\begin{aligned} \mathbb{P}\left(\boldsymbol{X}_{m+1} \in E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}, \widetilde{\boldsymbol{X}}_{m+1} \in R_{0,\boldsymbol{u}}\right) &\leq \mathbb{P}\left(\boldsymbol{X}_{m+1} \in E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}\right) \\ &\leq \int_{E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}} f_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x}, \end{aligned}$$

which is $O(\epsilon_m)$, because the volume of $E_{0,\boldsymbol{u},\boldsymbol{u}+\boldsymbol{\epsilon}_m}$ is $O(\epsilon_m)$ and the d-dimensional density $f_{\boldsymbol{X}}$ of the perpetuity is bounded on the state space $\mathbb{R}_+^d$. Yet, one could adopt a more general approach, and take the whole $\mathbb{R}^d$ as the state space. Then one could

consider the set of d-dimensional ϵ_m-strips defined as

$$\mathcal{S}(\epsilon_m) \;=\; \big\{\mathcal{E} \in \mathcal{B}\big(\mathbb{R}^d\big) : \mathbb{P}(\boldsymbol{X} \in \mathcal{E}) = O(\epsilon_m)\big\},$$

and then $\sup_{\mathcal{E}\in\mathcal{S}(\epsilon_m)} \mathbb{P}(\boldsymbol{X} \in \mathcal{E})$ could be considered as a possible bound. Yet, this would force us to reconsider the definition of exceedances, and examine whether, for instance, the set

$$]-\infty, u_1+\epsilon_m] \times \cdots \times]-\infty, u_d+\epsilon_m] \;\setminus\;]-\infty, u_1] \times \cdots \times]-\infty, u_d]$$

is in $\mathcal{S}(\epsilon_m)$. This could perhaps be achieved with regular variation considerations.

Chapter 9

Overall approximation of perpetuity exceedances

This is where we finally bring the previous results together. The parameters of the approximating compound Poisson model take the structure of the underlying perpetuity into account. Most importantly, the approximation is also carried out in the multidimensional case. So far, we have managed to link the exceedances of one-dimensional and d-dimensional perpetuities to relevant, $\{0,1\}$-valued, m-dependent processes. In order to put the final touch to the whole approximation, we now need to relate a finite sequence from a $\{0,1\}$-valued, m-dependent process to a compound Poisson measure. For this purpose, let us adopt the same approach as in Chapters 5 and 6, inspired by Barbour *et al.* [7].

9.1 Link between a $\{0,1\}$-valued, m-dependent process and its related compound Poisson measure

Let us consider a strictly stationary, $\{0,1\}$-valued, m-dependent process $(I_t)_{1\leq t\leq n}$, and use the same notation as in the Markovian cases of Chapters 5 and 6. Let $p=\mathbb{P}(I_1=1)$,

$$N_n(\bullet)=\sum_{i=1}^{n} I_i\delta_{ip}(\bullet)$$

and

$$\forall\, i\in\left\{1,\ldots,\left\lfloor\frac{n}{r}\right\rfloor\right\},\qquad T_{r,i}^{(m_1)}=\sum_{t=(i-1)r+1}^{ir-m_1} I_t\qquad T_{r,i}=\sum_{t=(i-1)r+1}^{ir} I_t.$$

We shall relate $N_n(\bullet)$ to the following compound Poisson measure. Still denoting by $\theta_r=\frac{1}{rp}\,\mathbb{P}\left(T_{r,1}\geq 1\right)$, $\nu_r=\mathcal{L}\left(T_{r,1}\middle|\, T_{r,1}\geq 1\right)$ and $\boldsymbol{\mu}$ the Lebesgue measure, the approximating compound Poisson process we use is

$$\boxed{\mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)}.$$

Since an m-dependent process forms an m-th order Markov chain, we can use the same Markovian notation as in Chapter 6. If $\forall\, i\in[\![1,m]\!], k_i\in\mathbb{N}$, and $\boldsymbol{k}\in\mathbb{N}^m$ such that $\boldsymbol{k}=(k_1,\ldots,k_m)$, we have

$$p_{\boldsymbol{k}} \;=\; \mathbb{P}\left(I_t=1\,|\,I_{t-k_1}=0,\ldots,I_{t-k_1-\cdots-k_m}=0\right).$$

We still denote by $\boldsymbol{v}_i\in\mathbb{N}^m$ the vector of indices such that $\boldsymbol{v}_i=\left(\mathbb{I}_{\{i\geq 1\}},\ldots,\mathbb{I}_{\{i\geq m\}}\right)$, and $\boldsymbol{w}_i\in\mathbb{N}^m$ such that

$$\boldsymbol{w}_i=(1,\ldots,1,r-m_1,1,\ldots,1).$$
$$\uparrow$$
$$i\text{-th position}$$

Theorem 9.1.1 *Let $(I_t)_{1\leq t\leq n}$ be a strictly stationary, $\{0,1\}$-valued, m-dependent process. Let $m_1\in\mathbb{N}^*$ such that $m_1\geq m$. Let $r\in\mathbb{N}^*$ such that $r\geq m+m_1$.*

If $\boldsymbol{\mu}$ denotes the Lebesgue measure on $[0,1]$, and $N_n(\bullet)$ the point process on $[0,1]$ introduced above, given the definitions for θ_r and ν_r, then

$$\begin{aligned} & d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), CP\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\ \leq\ & 2rp + 1 - (1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)\lfloor \frac{n}{r}\rfloor}\left(\prod_{j=1}^{m}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor \frac{n}{r}\rfloor - 1} \\ & + 1 - \left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor \frac{n}{r}\rfloor} + q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right), \end{aligned}$$

where $q_r = 1 - (1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(r-m)}$.

Proof. As in the Markovian case, let us create two i.i.d. auxiliary processes. For $m_1 \in \mathbb{N}^*$, let $\big((\widehat{T}_{r,i}, \widehat{T}_{r,i}^{(m_1)}\big)_{1\leq i\leq \lfloor n/r\rfloor}$ be i.i.d. replicas of $(T_{r,1}, T_{r,1}^{(m_1)})$. The triangular inequality we are going to use is the following.

$$\begin{aligned} & d_W\Big(\mathcal{L}\big(N_n(\bullet)\big), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\Big) \\ \leq\ & d_W\left(\mathcal{L}\big(N_n(\bullet)\big), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right)\right) && (9.1) \\ & + d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right) && (9.2) \\ & + d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right) && (9.3) \\ & + d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right)\right) && (9.4) \\ & + d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathrm{CP}\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\right) && (9.5) \end{aligned}$$

We can now evaluate the summands of the triangular inequality. The bound on (9.1) is given by the distortion argument, and by taking the last, incomplete block into account, giving

$$d_W\left(\mathcal{L}\big(N_n(\bullet)\big), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right)\right) \leq 2rp.$$

The bound on (9.2) is provided by Lemma 6.1.1, and gives

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right)$$
$$\leq \; 1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)\lfloor\frac{n}{r}\rfloor}\left(\prod_{j=1}^{m}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor\frac{n}{r}\rfloor-1}.$$

Here, we can note that, in the case when $r-m_1 > m$, we have $p_{\boldsymbol{w}_1} = p$ and $p_{\boldsymbol{w}_j} = p_{\boldsymbol{v}_{j-1}}$ for $j \in [\![2, m]\!]$, because we are dealing with an m-dependent process, and consequently jumping above a big block erases the memory. In that case, this part of the bound becomes

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right)$$
$$\leq \; 1-\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}.$$

Now, focusing on (9.3), let us recall that after using Lindeberg's [48] method of decomposition and Berbee's [14] Lemma, the following bound on (9.3) can be found in Barbour *et al.* [7]:

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right) \;\leq\; \frac{n}{r}\beta_I(m_1),$$

where β_I is the absolute regularity coefficient of the sequence $(I_t)_{t\in\mathbb{N}^*}$. Given that $(I_t)_{t\in\mathbb{N}^*}$ is m-dependent, it follows that $\beta_I(m_1) = 0$ for $m_1 \geq m$, and consequently the distance (9.3) vanishes in such a case. The bound on (9.4) is provided by Lemma 6.1.2, and gives

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} T_{r,i}^{(m_1)}\delta_{irp}(\bullet)\right)\right)$$
$$\leq \; 1-\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}.$$

The fourth and final bound stems from Xia's result (*c.f.* [68]).

$$d_W\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r\rfloor} \widehat{T}_{r,i}\delta_{irp}(\bullet)\right), \mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet), \nu_r\right)\right) \;\leq\; q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right),$$

where $q_r = 1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(r-m)}$. This completes the proof. □

9.2 Complete approximation

Finally, and this is the main result of the present work, we get an overall bound on the distance between the exceedances of our perpetuity and a judicious compound Poisson measure.

9.2.1 One-dimensional setting

In a first step, we notice that the perpetuity generated by $X_t = B_t X_{t-1} + A_t$ satisfies the following relation:

$$X_t = \prod_{j=0}^{m-1} B_{t-j} X_{t-m} + \sum_{k=t-m+1}^{t} A_k \prod_{j=k+1}^{t} B_j,$$

and therefore, under some mixing conditions, should not be too different from the process generated by

$$\widetilde{X}_t = \prod^{m-1} B_{t-j} \widehat{X}_{t-m} + \sum^{t} A_k \prod^{t} B_j,$$

where $(\widehat{X}_t)$ are independent copies of X_1. This latter process has the advantage of being m-dependent, and so has its associated process of threshold exceedances. So the first step, made in Chapter 7 for the one-dimensional case, and in Chapter 8 for the multidimensional case, is the approximation of the exceedance process of (X_t) by the exceedance process of $(\widetilde{X}_t)$. The second step was made in the previous section, where $\{0,1\}$-valued, m-dependent processes are linked to a compound Poisson measure. Let us write down the whole result more formally, in the one-dimensional case. Here, we are denoting $p = \mathbb{P}(X_1 > u)$ and, for each $m \in \mathbb{N}^*$, for $\boldsymbol{k} = (k_1, \ldots, k_m) \in \mathbb{N}^m$, $p_{\boldsymbol{k}} = \mathbb{P}\left(\widetilde{X}_t > u \,\middle|\, \widetilde{X}_{t-k_1} \leq u, \ldots, \widetilde{X}_{t-k_1-\cdots-k_m} \leq u\right)$ and $\widetilde{X}_s = \prod_{j=0}^{m-1} B_{s-j} \widehat{X}_{s-m} + \sum_{k=s-m+1}^{s} A_k \prod_{j=k+1}^{s} B_j$, for $s > m$, where $\widehat{X}_{s-m}$ is an independent copy of X_1. For $1 \leq s \leq m$, $\widetilde{X}_s = X_s$. As for the parameters of the approximating compound Poisson measure, the intensity is determined by $\theta_r = \frac{q_r}{rp}$, where $q_r = 1 - (1-p)\prod_{i=1}^{m-1}(1 -$

$p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(r-m)}$, and the compounding distribution $\nu_r = \mathcal{L}(T_{r,1}|T_{r,1} \geq 1)$, where $T_{r,1} = \sum_{t=1}^{r} \mathbb{I}_{\{\tilde{X}_t > u\}}$. An explicit form is given by Equation (6.2) in Section 6.2.

Theorem 9.2.1 *Let $(X_t)_{1\leq t\leq n}$ be a random process on $\mathbb{R}$ satisfying the stochastic recursion specified in Equation (7.1). Assume that $\mathbb{E}\ \ln^+|A| < +\infty$, $-\infty \leq \mathbb{E}\ \ln|B| < 0$, and X_1 has the stationary distribution of Theorem 7.2.1, admits a bounded density function f_X, and has finite second moment $\mathbb{E}X^2$. Then, for all m in $[\![1,n]\!]$ such that $2m < n$, for all r in $[\![2m,n]\!]$, given the notation above, we have*

$$
\begin{aligned}
& d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), CP\big(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\big)\right) \\
&\leq\ 8(n-m)\big\|f_X\big\|_\infty^{2/3}\big(\mathbb{E}X^2\big)^{1/3}\big(\mathbb{E}B^2\big)^{m/3} \\
&\quad +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),
\end{aligned}
$$

where $q_r = 1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(r-m)}$.

Remark. Let us look at conditions for the right-hand side to tend to zero as $n \to \infty$, *c.f.* Barbour *et al.* [7]. For the first term of the bound to tend to zero, we need $\mathbb{E}B^2 < 1$. Then, this part tends to zero for m-dependence parameters $m_n = \lfloor n^\alpha \rfloor$, where $0 < \alpha < 1$. Let us focus now on the second term. For a threshold $u = u_n$ such that $n\mathbb{P}(X > u_n) \to t \in]0,+\infty[$, for a sequence $m_n \to +\infty$ such that $m_n/\beta(m_n) \sim n/t$, and a sequence of block sizes $r_n = \lfloor\sqrt{\beta(m_n)/t}\rfloor$, the second part of the bound is of order $O\big(\sqrt{t\beta(m_n)}\big)$, which tends to zero if our perpetuity is absolutely regular. This is the case whenever (X_t) is ψ-irreducible, and there is $\epsilon > 0$ such that $\mathbb{E}|B|^\epsilon < 1$ and $\mathbb{E}|A|^\epsilon < +\infty$, *c.f.* Basrak *et al.* [12].

Proof of Theorem 9.2.1. Adding the bounds given in Theorem 7.4.1 of Chapter 7 and Theorem 9.1.1 of the last section, we obtain, for all integers $m_1 \geq m$,

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\big(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\big)\right) \\ \leq\ & 8(n-m)\big\|f_X\big\|_\infty^{2/3}\big(\mathbb{E}X^2\big)^{1/3}\big(\mathbb{E}B^2\big)^{m/3} \\ & +2rp+1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)\lfloor\frac{n}{r}\rfloor}\left(\prod_{j=1}^{m}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor\frac{n}{r}\rfloor-1} \\ & +1-\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right), \end{aligned}$$

which, in the case when $r-m_1 > m$, simplifies into

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\big(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\big)\right) \\ \leq\ & 8(n-m)\big\|f_X\big\|_\infty^{2/3}\big(\mathbb{E}X^2\big)^{1/3}\big(\mathbb{E}B^2\big)^{m/3} \\ & +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right), \end{aligned}$$

and this bound is minimised over the choices of m_1 by putting $m_1 = m$, which finally gives

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\big(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\big)\right) \\ \leq\ & 8(n-m)\big\|f_X\big\|_\infty^{2/3}\big(\mathbb{E}X^2\big)^{1/3}\big(\mathbb{E}B^2\big)^{m/3} \\ & +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right). \end{aligned}$$

Hence the result. □

We shall later refer to the various elements of the bound in the numerical illustrations of Chapter 10 for the one-dimensional case as

$$\begin{aligned} M_n(1) &= 8(n-m)\|f_X\|_\infty^{2/3}\left(\mathbb{E}X^2\right)^{1/3}\left(\mathbb{E}B^2\right)^{m/3}, \\ M_n(2) &= 2rp, \\ M_n(3) &= 2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor \frac{n}{r} \rfloor}, \\ M_n(4) &= q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right), \end{aligned}$$

and to the overall bound as $M_n = M_n(1) + M_n(2) + M_n(3) + M_n(4)$.

9.2.2 Multidimensional setting

In the multidimensional framework, the same approach is adopted. Yet, this time, we only look at non-negative situations, *i.e.* the state space is $\mathbb{R}_+^d$. We are considering vectors $(\boldsymbol{X}_t)_{1\le t\le n}$ generated by the random recurrence $\boldsymbol{X}_t = \boldsymbol{B}_t\boldsymbol{X}_{t-1} + \boldsymbol{A}_t$, where $(\boldsymbol{B}_t)$ are $d \times d$ random matrices with coefficients in $\mathbb{R}_+$, and $(\boldsymbol{A}_t)$ are d-dimensional random vectors, also with real, non-negative coefficients, such that $\big((\boldsymbol{A}_t, \boldsymbol{B}_t)\big)$ are jointly i.i.d. Again, we can rewrite the multidimensional perpetuity as

$$\boldsymbol{X}_t = \prod_{j=0}^{m-1}\boldsymbol{B}_{t-j}\boldsymbol{X}_{t-m} + \sum_{k=t-m+1}^{t}\prod_{j=k+1}^{t}\boldsymbol{B}_j\boldsymbol{A}_k.$$

Then, the multivariate process generated by

$$\widetilde{\boldsymbol{X}}_t = \prod_{j=0}^{m-1}\boldsymbol{B}_{t-j}\widehat{\boldsymbol{X}}_{t-m} + \sum_{k=t-m+1}^{t}\prod_{j=k+1}^{t}\boldsymbol{B}_j\boldsymbol{A}_k,$$

where $(\widehat{\boldsymbol{X}}_t)$ are independent copies of $\boldsymbol{X}_1$, is m-dependent. This time, given that the state space is $\mathbb{R}_+^d$, we are interested in exceedances out of volumes of the type $R_{0,\boldsymbol{u}} =]0, u_1] \times \cdots \times]0, u_d]$. In Chapter 8, we approximated the exceedance process of $(\boldsymbol{X}_t)$ by the m-dependent one of $(\widetilde{\boldsymbol{X}}_t)$, and the compound Poisson approximation

for the latter is provided in the last section. In this new setting, we are now denoting $p = \mathbb{P}(\boldsymbol{X}_1 \notin R_{0,\boldsymbol{u}})$ and, for each $m \in \mathbb{N}^*$, for $\boldsymbol{k} = (k_1, \ldots, k_m) \in \mathbb{N}^m$, $p_{\boldsymbol{k}} = \mathbb{P}\left(\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}} \,\middle|\, \widetilde{\boldsymbol{X}}_{t-k_1} \in R_{0,\boldsymbol{u}}, \ldots, \widetilde{\boldsymbol{X}}_{t-k_1-\cdots-k_m} \in R_{0,\boldsymbol{u}}\right)$ and $\widetilde{\boldsymbol{X}}_s = \prod_{j=0}^{m-1} \boldsymbol{B}_{s-j} \widehat{\boldsymbol{X}}_{s-m} + \sum_{k=s-m+1}^{s} \prod_{j=k+1}^{s} \boldsymbol{B}_j \boldsymbol{A}_k$, for $s > m$, where $\widehat{\boldsymbol{X}}_{s-m}$ is an independent copy of $\boldsymbol{X}_1$. For $1 \leq s \leq m$, $\widetilde{\boldsymbol{X}}_s = \boldsymbol{X}_s$. As for the parameters of the approximating compound Poisson measure, we still have $\theta_r = \frac{q_r}{rp}$ and $\nu_r = \mathcal{L}(T_{r,1} | T_{r,1} \geq 1)$, with the new definitions of p and $q_r = 1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(r-m)}$, noting that now $T_{r,1} = \sum_{t=1}^{r} \mathbb{I}_{\{\widetilde{\boldsymbol{X}}_t \notin R_{0,\boldsymbol{u}}\}}$.

Theorem 9.2.2 *Let us assume that the process $(\boldsymbol{X}_t)_{1 \leq t \leq n}$ follows Equation (8.1), with random coefficients $(\boldsymbol{A}_t, \boldsymbol{B}_t)$ which have non-negative entries, that $\mathbb{E} \ln^+ |||\boldsymbol{B}_1||| < +\infty$, $\mathbb{E} \ln^+ ||\boldsymbol{A}_1||_2 < +\infty$ and $\mathbb{E} \ln |||\boldsymbol{B}_1||| < 0$, which ensure its strict stationarity. Let $\boldsymbol{X}$ be a random variable on $\mathbb{R}_+^d$ which has the stationary distribution of the perpetuity, as defined in Theorem 8.2.1. Let us assume that $\mathbb{E}||\boldsymbol{X}||_2^2 < +\infty$, and that $\boldsymbol{X}$ has a bounded density function $f_{\boldsymbol{X}}$ on the state space $S = \mathbb{R}_+^d$. Then, for every m in $[\![1, n]\!]$ such that $2m < n$, for all $\epsilon_m > 0$, for all r in $[\![2m, n]\!]$, given the notation above, we have*

$$d_W\left(\mathcal{L}\left(\sum_{t=1}^{n} \mathbb{I}_{\{\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}}\}} \delta_{tp}(\bullet)\right), CP\big(\theta_r \boldsymbol{\mu}(\bullet), \nu_r\big)\right)$$

$$\leq \frac{8(n-m)}{\epsilon_m^2}\Big(\mathbb{E}|||\boldsymbol{B}|||^2\Big)^m \mathbb{E}||\boldsymbol{X}||_2^2 + 2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m) - \prod_{i=1}^{d} u_i\right)\|f_{\boldsymbol{X}}\|_\infty$$

$$+2rp + 2 - 2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor \frac{n}{r} \rfloor} + q_r\left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\right).$$

Remark. Again, let us find conditions for the right-hand side to tend to zero, *c.f.* Barbour *et al.* [7]. For the convergence of the first term, we need $\mathbb{E}|||\boldsymbol{B}|||^2 < 1$. Then, we have convergence to zero for m-dependence parameters $m_n = \lfloor n^\alpha \rfloor$, where $0 < \alpha < 1$. Concerning the second term, the same conditions as in the one-dimensional case are sufficient. For a threshold $\boldsymbol{u} = \boldsymbol{u}_n$ such that $n\mathbb{P}(\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}_n}) \to t \in]0, +\infty[$, for a sequence $m_n \to +\infty$ such that $m_n/\beta(m_n) \sim n/t$, and a sequence of block sizes $r_n = \lfloor \sqrt{\beta(m_n)/t} \rfloor$, the second part of the bound is of order $O\big(\sqrt{t\beta(m_n)}\big)$, which tends to zero if our perpetuity is absolutely regular. This happens whenever $(\boldsymbol{X}_t)$

is ψ-irreducible, and there is $\epsilon > 0$ such that $\mathbb{E}\|\boldsymbol{B}\|^{\epsilon} < 1$ and $\mathbb{E}\|\boldsymbol{A}\|_2^{\epsilon} < +\infty$, *c.f.* Basrak *et al.* [12].

Proof of Theorem 9.2.2. Let us add the bounds given in Theorem 8.5.1 of Chapter 8 and Theorem 9.1.1 of the last section, and we obtain, for all integers $m_1 \geq m$,

$$\begin{aligned}
&d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right)\\
&\leq \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_{\infty}\\
&\quad +2rp+1-(1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)\lfloor\frac{n}{r}\rfloor}\left(\prod_{j=1}^{m}(1-p_{\boldsymbol{w}_j})\right)^{\lfloor\frac{n}{r}\rfloor-1}\\
&\quad +1-\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),
\end{aligned}$$

which, in the case when $r-m_1 > m$, simplifies into

$$\begin{aligned}
&d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right)\\
&\leq \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_{\infty}\\
&\quad +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})(1-p_{\boldsymbol{v}_m})^{(m_1-m)}\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),
\end{aligned}$$

and again, this is minimised over the choices of m_1 by putting $m_1 = m$, which gives

$$\begin{aligned}
&d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right)\\
&\leq \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2 + 2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\|f_{\boldsymbol{X}}\|_{\infty}\\
&\quad +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),
\end{aligned}$$

and completes the proof. □

We shall later refer to the various elements of the bound in the numerical illustrations of Chapter 10 for the multidimensional case as

$$\begin{aligned}
\mathcal{M}_n(1) &= \frac{8(n-m)}{\epsilon_m^2}\Big(\mathbb{E}\|\boldsymbol{B}\|^2\Big)^m \mathbb{E}\|\boldsymbol{X}\|_2^2, \\
\mathcal{M}_n(2) &= 2(n-m)\Big(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\Big)\|f_{\boldsymbol{X}}\|_\infty, \\
\mathcal{M}_n(3) &= 2rp, \\
\mathcal{M}_n(4) &= 2-2\Big((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\Big)^{\lfloor \frac{n}{r} \rfloor}, \\
\mathcal{M}_n(5) &= q_r\Big(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\Big),
\end{aligned}$$

and to the overall bound as $\mathcal{M}_n = \mathcal{M}_n(1) + \mathcal{M}_n(2) + \mathcal{M}_n(3) + \mathcal{M}_n(4) + \mathcal{M}_n(5)$.

As indicated in Section 3.3, a compound Poisson approximation of multivariate exceedances can also be found in Basrak *et al.* [12], for the asymptotic case. The type of exceedance considered there is the same, *i.e.* indicators telling whether a point is in or out of a limit rectangle. Yet the authors work in the asymptotic framework, where sequences are infinite, and the parameters of the approximating compound Poisson measure are infinite sums. This present work is about the finite case, where finite sequences are modelled with empirical parameters. In particular, we can choose these so that the related m and r are optimal, in the sense that they minimise the bound, for a given sequence of length n.

Chapter 10

Implementation for the calculation of Value-at-Risk

We are now going to use the models developed in the previous chapters to illustrate in which manner the Value-at Risk figure is exceeded by a perpetuity. We shall look at the same GARCH(1,1) model that was selected earlier in Chapter 5 to describe the evolution of the log-returns of Sterling versus US Dollar exchange rates. Let us recall its specifications. The log-returns $(X_t)_{t\in\mathbb{N}^*}$ are modelled as follows, using a sequence $(Z_t)_{t\in\mathbb{N}^*}$ of i.i.d. symmetric random variables, called innovations, and a sequence $(\sigma_t)_{t\in\mathbb{N}^*}$ of random variables, called volatilities, such that

$$\forall\, t\in\mathbb{N}^*, \quad X_t = \sigma_t Z_t, \tag{10.1}$$

where σ_t verifies

$$\forall\, t\in\mathbb{N}^*, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{10.2}$$

which can be rewritten as

$$\forall t \in \mathbb{N}^*, \quad \sigma_t^2 = \alpha_0 + \left(\alpha_1 Z_{t-1}^2 + \beta_1\right)\sigma_{t-1}^2, \tag{10.3}$$

and the quantity σ_t^2, called the *conditional variance* or *volatility*, clearly forms a perpetuity, given that the equation above falls into our definition of a stochastic recursion, with $A_t = \alpha_0$ and $B_t = \alpha_1 Z_{t-1}^2 + \beta_1$. We shall use the estimates of the parameters from Chapter 4, and set the number of degrees of freedom to be 12 instead of 6.84 in order to accelerate the convergence:

$$\alpha_0 = 1.331 \times 10^{-7}$$

$$\alpha_1 = 2.482 \times 10^{-2}$$

$$\beta_1 = 9.573 \times 10^{-1}.$$

This is a plot of a realisation of the resulting perpetuity. We are looking here at the volatility of the GARCH(1,1) process, not at the GARCH(1,1) process itself.

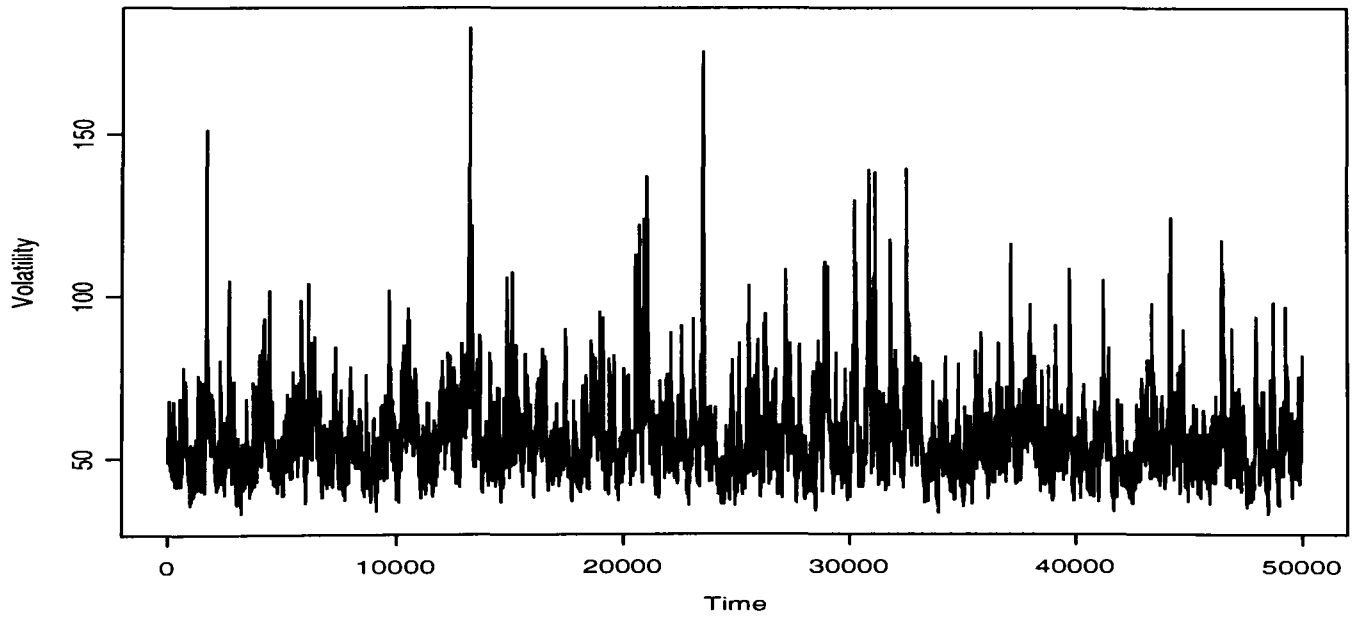

Figure 10.1: *Sequence of 50,000 points from the volatility of the GARCH(1,1) model specified above.*

10.1 Threshold u: the Value-at-Risk

This work has been concerned with financial time series, and the rate at which some of them exceed thresholds. The compound Poisson model gives a direct way to access the regulatory risk measure. Recall that, according to the Basel Accords, the Value-at-Risk associated with a given security is three times the value (of a loss, or a change in returns) that has exactly 1% chance to be exceeded in ten days. Here, we approximate the empirical point process of exceedances $\sum_{i=1}^{n} \mathbb{I}_{\{X_i>u\}}\delta_{ip}(\bullet)$ on the measurable space $[0, np]$ by the compound Poisson random measure $\mathrm{CP}(\theta_r \boldsymbol{\mu}(\bullet), \nu_r)$. In the present case of exchange rates, we are dealing with daily data, and hence n is the number of days. There are two ways we can use the approximation developed in this thesis. The first approach, carried out in Section 10.2, is to consider a sequence of length $n = 10$, and look at how often a threshold u_{10} is exceeded. Denoting $p_{10} = \mathbb{P}(X_1 > u_{10})$, Theorem 9.2.1 in Chapter 9 provides an approximation on the measurable space $[0, 10p_{10}]$ of the form

$$d_W\left(\mathcal{L}\Big(\sum_{i=1}^{10} \mathbb{I}_{\{X_i>u_{10}\}}\delta_{ip_{10}}(\bullet)\Big), \mathrm{CP}\big(\theta_{r_{10}}\boldsymbol{\mu}(\bullet), \nu_{r_{10}}\big)\right) \;\leq\; M_{10},$$

and given that $\boldsymbol{\mu}\big([0, 10p_{10}]\big) = 10p_{10}$, the properties of the Wasserstein distance yield

$$\left|\mathbb{E}\Big(\sum_{i=1}^{10} \mathbb{I}_{\{X_i>u_{10}\}}\Big) - \mathbb{E}\Big(\mathrm{CP}\big(10p_{10}\theta_{r_{10}}, \nu_{r_{10}}\big)\Big)\right| \;\leq\; 10M_{10},$$

which provides us with a bound on the approximation error. Now we can also look at longer sequences of size n, as will be the case in Section 10.3, on a measurable space $[0, np_n]$, with $p_n = \mathbb{P}(X_1 > u_n)$, and this would translate as

$$d_W\left(\mathcal{L}\Big(\sum_{i=1}^{n} \mathbb{I}_{\{X_i>u_n\}}\delta_{ip_n}(\bullet)\Big), \mathrm{CP}\big(\theta_{r_n}\boldsymbol{\mu}(\bullet), \nu_{r_n}\big)\right) \;\leq\; M_n.$$

After having chosen the threshold u_n of interest, such that $p_n = \mathbb{P}(X_1 > u_n)$, we obtain the approximation bound M_n. But then, given the nature of point processes,

this M_n bound is valid for any element of $\mathcal{B}\big([0, np_n]\big)$, and in particular for $[0, 10p_n]$, which corresponds to a count over ten consecutive days. But we could also be looking at other sets of time points such as, say, the set of all Mondays, and M_n would still bound the corresponding compound Poisson approximation. The drawback of this potentially more powerful approach, encompassing more configurations, is that for large n and fixed p_{10}, M_n will be larger than M_{10}, and possibly useless, given that a distance between distributions is always smaller than 1.

Now, in order for the convergence to occur, according to the Remark following Theorem 9.2.1, we must have $np_n \to t$, for some t in $\mathbb{R}_+$. So the validity of the approximation can be checked using small and large values of n, say n_1 and n_2 respectively, and corresponding thresholds u_{n_1} and u_{n_2} such that $n_1\mathbb{P}(X_1 > u_{n_1}) = n_2\mathbb{P}(X_1 > u_{n_2})$. In our case, we shall use $n_1 = 10$ because we want to look at the ten-day VaR, and $n_2 = 10,000$. Now p_{10} is determined by the VaR confidence level. Indeed, we are looking for the threshold that makes at least one exceedance occur with a probability of 1% in ten days. Hence, if our exceedance process were exactly behaving like its compound Poisson approximation, we would need to find u_{10} such that $\mathbb{E}\big(\mathrm{CP}(10p_{10}\theta_{r_{10}}, \nu_{r_{10}})\big) = 0.01$, in other words, such that there are on average 0.01 exceedances every ten days. Now,

$$\begin{aligned} \mathbb{E}\big(\mathrm{CP}(10p_{10}\theta_{r_{10}}, \nu_{r_{10}})\big) = 0.01 \quad &\Longleftrightarrow \quad 10p_{10}\theta_{r_{10}}\mathbb{E}\nu_{r_{10}} = 0.01 \\ &\Longleftrightarrow \quad \frac{\mathbb{P}\big(T_{r_{10},1} \geq 1\big)}{r_{10}p_{10}} \times 10p_{10} \times \frac{\mathbb{E}T_{r_{10},1}}{\mathbb{P}\big(T_{r_{10},1} \geq 1\big)} = 0.01 \\ &\Longleftrightarrow \quad p_{10} = 0.001, \end{aligned} \tag{10.4}$$

because $\mathbb{E}T_{r,1} = rp$, and hence $10p_{10} = 0.01$. Therefore we can later look at the convergence to compound Poisson by constraining $np_n = 0.01$, thus choosing u_n such that $\mathbb{P}(X_1 > u_n) = 0.01/n$. This will be done in Section 10.2. Now Equation (10.4) tells us that the threshold we are looking at for the ten-day VaR is the 99.9% quantile of the perpetuity. Since we do not know the exact distribution of perpetuities, we shall use simulations to obtain a precise estimate of the VaR for the perpetuity we

have chosen. The confidence interval quoted below is therefore totally unrelated to any dataset. Running 100 simulations of sequences of length 1,000,000, we obtain an estimate and a 99% confidence interval for the VaR, using the central limit theorem: if $(X_i)_{1\leq i\leq n}$ are i.i.d. random variables with mean μ, then $\sqrt{n}(\overline{X}-\mu)/s \to N(0,1)$ as n tends to infinity, in distribution, where $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i-\overline{X})$. Consequently, an approximate 99% confidence interval for the sample mean is given by the quantiles of the standard normal distribution, based on a sample of length $n = 100$:

$$\overline{X} \pm \frac{2.58}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n-1}}.$$

In the case of the GARCH(1,1) process from the previous section, this gives an estimate of VaR=129.41 with approximate confidence interval of [129.39, 129.44]. This 99.9% quantile is the ten-day VaR given by the compound Poisson viewpoint. Now, we do not know for sure that our empirical point process of exceedances behaves in such a manner. So far, our theoretical work shows that there is a bound on the approximation, even though at this stage we are not sure how good it is. Or, putting the argument differently, there are more reasons to act as if the empirical exceedance process were compound Poisson than as if it were i.i.d. Gaussian, which is the attitude commonly adopted in the banking literature. Some recent works, *c.f.* Borkovec and Klüppelberg [17] and Longin [50], have used extreme value theory to estimate the VaR, *c.f.* Formula (28) in Borkovec and Klüppelberg [17], and the latter illustrate the fact that Gaussian assumptions lead to an underestimation of quantiles, and hence of the Value-at-Risk, especially when little data is available. In our present case, we are dealing with simulated data, and can therefore safely use the empirical quantiles from very long sequences, and concentrate on other issues, but one seldom has such an opportunity when dealing with real data. Our goal here is to illustrate the method and discuss the quality of the theoretical bounds. This quality will be assessed in Section 10.2, where short sequences are used, whereas the compound Poisson behaviour will be described for longer sequences in Section 10.3.

10.2 Bound and convergence

In a first step we shall calculate the approximation bound for ten-day exceedances, corresponding to the ten-day VaR. Then, in order to check the convergence, we shall increase the length of the sequence to $n = 10,000$ and adjust the threshold so that $p_{10,000} = 0.01/10,000 = 10^{-6}$. This is carried out for our GARCH(1,1) process, and also for the perpetuity generated by $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, where $(Z_t)_{1\le t\le n}$ and $(\epsilon_t)_{1\le t\le n}$ are i.i.d., t_{10}-distributed sequences independent of each other. Such a perpetuity is indeed more ergodic than the GARCH(1,1) process, and therefore enables us to illustrate the present approximation in a more striking way. The overall bound in this one-dimensional case is the following.

$$\begin{aligned} & d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{X_t>u\}}\delta_{tp}(\bullet)\right), \mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right) \\ \le\ & 8(n-m)\left\|f_X\right\|_\infty^{2/3}\left(\mathbb{E}X^2\right)^{1/3}\left(\mathbb{E}B^2\right)^{m/3} \\ & +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right). \end{aligned}$$

As mentioned after the proof of Theorem 9.2.1, we denote

$$\begin{aligned} M_n(1) &= 8(n-m)\left\|f_X\right\|_\infty^{2/3}\left(\mathbb{E}X^2\right)^{1/3}\left(\mathbb{E}B^2\right)^{m/3} \\ M_n(2) &= 2rp \\ M_n(3) &= 2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor} \\ M_n(4) &= q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right), \end{aligned}$$

so that the Wasserstein distance is bounded by $M_n = M_n(1)+M_n(2)+M_n(3)+M_n(4)$.

10.2.1 Parameter estimation

First, we need to have a rough idea of how big M_n is. For this purpose, a sequence of length 2,000,000 is generated from the perpetuity specified above, and $\left\|f_X\right\|_\infty$ is

estimated with the default kernel density estimator of S-PLUS, which is

$$\hat{f}(x) = \frac{1}{b}\sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right).$$

The kernel K is the density function of the standard normal distribution, the bandwidth b is four times the standard deviation of the sample, and x_j are the points at which the density is estimated, *c.f.* Bowman and Azzalini [18] for this particular kernel density estimator, and Silverman [63] for density estimation in general. The supremum of the density is roughly estimated by the maximum of the density function obtained. Then, $\mathbb{E}X^2$ is estimated from the simulated sample. Now, we have an exact expression of $q_r = \mathbb{P}(T_{r,1} \geq 1)$, but it relies exclusively on the transition probabilities, which we can only estimate from simulated sequences, given that we only have asymptotic expressions for the distribution of a perpetuity. Unfortunately, because exceedances above the VaR threshold are rare events, there is little data to estimate the values of these transition probabilities robustly, which are very sensitive to the simulated sequence they are computed from. Consequently, in order to overcome this, we can directly estimate the end product by

$$\hat{q}_r = \text{(number of blocks with exceedances)/(total number of blocks)},$$

which varies much less across simulated samples. Let us look at its bias.

$$\hat{q}_r \quad = \quad \frac{1}{\lfloor n/r \rfloor}\sum_{i=1}^{\lfloor n/r \rfloor} \mathbb{I}_{\{T_{r,i} \geq 1\}},$$

consequently

$$\mathbb{E}\,\hat{q}_r = \frac{1}{\lfloor n/r \rfloor}\sum_{i=1}^{\lfloor n/r \rfloor} \mathbb{E}\,\mathbb{I}_{\{T_{r,i} \geq 1\}} = \frac{1}{\lfloor r/n \rfloor}\sum_{i=1}^{\lfloor n/r \rfloor} \mathbb{E}\,\mathbb{I}_{\{T_{r,1} \geq 1\}} = \mathbb{E}\,\mathbb{I}_{\{T_{r,1} \geq 1\}} = \mathbb{P}(T_{r,1} \geq 1),$$

since the process is strictly stationary. Hence, $\hat{q}_r$ is unbiased. Now, let us look at its asymptotic variance.

$$
\begin{aligned}
\operatorname{Var}\hat{q}_r &= \frac{1}{\lfloor n/r\rfloor^2}\mathbb{E}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\mathbb{I}_{\{T_{r,i}\geq 1\}} - \left\lfloor\frac{n}{r}\right\rfloor\mathbb{P}(T_{r,1}\geq 1)\right)^2 \\
&= \frac{1}{\lfloor n/r\rfloor^2}\mathbb{E}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}} - \mathbb{P}(T_{r,1}\geq 1)\Big)\right)^2 \\
&= \frac{1}{\lfloor n/r\rfloor^2}\left(\sum_{i=1}^{\lfloor n/r\rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,i}\geq 1\}} + 2\sum_{i<j}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,j}\geq 1\}}\Big)\right) \\
&= \frac{1}{\lfloor n/r\rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r\rfloor^2}\sum_{i<j}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,j}\geq 1\}}\Big) \\
&= \frac{1}{\lfloor n/r\rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r\rfloor^2}\sum_{i=1}^{\lfloor n/r\rfloor-1}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,i+1}\geq 1\}}\Big) \\
&\quad + \frac{2}{\lfloor n/r\rfloor^2}\sum_{j-i\geq 2}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,j}\geq 1\}}\Big) \\
&= \frac{1}{\lfloor n/r\rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r\rfloor}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,1}\geq 1\}},\mathbb{I}_{\{T_{r,2}\geq 1\}}\Big) \\
&\quad + \frac{2}{\lfloor n/r\rfloor^2}\sum_{j-i\geq 2}\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,j}\geq 1\}}\Big) \\
&\leq \frac{1}{\lfloor n/r\rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r\rfloor}\Big|\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,1}\geq 1\}},\mathbb{I}_{\{T_{r,2}\geq 1\}}\Big)\Big| \\
&\quad + \frac{2}{\lfloor n/r\rfloor^2}\sum_{j-i\geq 2}\Big|\operatorname{Cov}\Big(\mathbb{I}_{\{T_{r,i}\geq 1\}},\mathbb{I}_{\{T_{r,j}\geq 1\}}\Big)\Big|.
\end{aligned}
$$

Now, we know from Lemma 3 in Section 1.2.2 of Doukhan [27] that if X and Y are two random variables, then their covariance is bounded by the coefficient of absolute regularity between their σ-fields, *i.e.* $|\operatorname{Cov}(X,Y)| \leq \beta\big(\sigma(X),\sigma(Y)\big)$. Note that

$$
\begin{aligned}
&\beta\Big(\sigma\big(\mathbb{I}_{\{T_{r,i}\geq 1\}}\big),\sigma\big(\mathbb{I}_{\{T_{r,j}\geq 1\}}\big)\Big) \\
&\quad\leq \beta\Big(\sigma\big(\{X_{(i-1)r+1}>u\},\ldots,\{X_{ir}>u\}\big),\sigma\big(\{X_{(j-1)r+1}>u\},\ldots,\{X_{jr}>u\}\big)\Big) \\
&\quad\leq \beta_X\big((|i-j|-1)r\big),
\end{aligned}
$$

where β_X is the β-mixing coefficient of the underlying process. As our process of squared volatilities satisfies the two conditions $\mathbb{E}B<1$ and $\mathbb{E}A<+\infty$, we can infer

from Basrak *et al.* [12] that it is geometrically ergodic, and hence geometrically absolutely regular, therefore there exist $C > 0$ and $\rho \in]0,1[$ such that $\beta_X(m) \leq C\rho^m$. This implies that

$$\begin{aligned}
\operatorname{Var}\hat{q}_r &\leq \frac{1}{\lfloor n/r \rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r \rfloor}\left|\operatorname{Cov}\left(\mathbb{I}_{\{T_{r,1}\geq 1\}}, \mathbb{I}_{\{T_{r,2}\geq 1\}}\right)\right| \\
&\quad + \frac{2}{\lfloor n/r \rfloor^2}\sum_{j-i\geq 2} C\rho^{|i-j|r-r} \\
&\leq \frac{1}{\lfloor n/r \rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r \rfloor}\left|\operatorname{Cov}\left(\mathbb{I}_{\{T_{r,1}\geq 1\}}, \mathbb{I}_{\{T_{r,2}\geq 1\}}\right)\right| \\
&\quad + \frac{2C}{\lfloor n/r \rfloor^2}\sum_{k=2}^{\lfloor n/r \rfloor}\left(\left\lfloor \frac{n}{r} \right\rfloor - k\right)\rho^{kr-r} \\
&\leq \frac{1}{\lfloor n/r \rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r \rfloor}\left|\operatorname{Cov}\left(\mathbb{I}_{\{T_{r,1}\geq 1\}}, \mathbb{I}_{\{T_{r,2}\geq 1\}}\right)\right| + \frac{2C}{\lfloor n/r \rfloor}\sum_{k=1}^{\lfloor n/r \rfloor - 1}\rho^{kr} \\
&\leq \frac{1}{\lfloor n/r \rfloor}\operatorname{Var}\mathbb{I}_{\{T_{r,1}\geq 1\}} + \frac{2}{\lfloor n/r \rfloor}\left|\operatorname{Cov}\left(\mathbb{I}_{\{T_{r,1}\geq 1\}}, \mathbb{I}_{\{T_{r,2}\geq 1\}}\right)\right| + \frac{2C(\rho^r - \rho^{r\lfloor n/r \rfloor})}{\lfloor n/r \rfloor(1-\rho^r)}.
\end{aligned}$$

Given that we are looking at sequences of block sizes such that $n/r_n \to +\infty$, *c.f.* Barbour *et al.* [7], the right-hand side of the equation tends to zero as n tends to infinity. This implies that the asymptotic variance of $\hat{q}_r$ is zero. In particular, given that $\hat{q}_r$ is also unbiased, $\hat{q}_r$ is a weakly consistent estimator of q_r.

Now, in the special case of the GARCH(1,1) process, from Equation (10.3), we have $B_t = \alpha_1 Z_t^2 + \beta_1$. According to our assumptions, we know that Z follows a Student distribution with 12 degrees of freedom, hence the expectation of B_t^2 is easily obtained. Indeed, $\mathbb{E}B_t^2 = \alpha_1^2\mathbb{E}Z_t^4 + 2\alpha_1\beta_1\mathbb{E}Z_t^2 + \beta_1^2$, and the second and fourth moments of a Student distribution with d degrees of freedom ($d>4$) are given by

$$\mathbb{E}Z_t^2 = \frac{d}{d-2} \qquad \text{and} \qquad \mathbb{E}Z_t^4 = \frac{3d^2}{(d-4)(d-2)}.$$

Hence we get $\mathbb{E}B_t^2 = 0.9768$. Now, this is strictly smaller than 1, which ensures both stationarity and convergence of the exceedance process to the compound Poisson measure. Yet, this exponent is still fairly close to 1, which means that the convergence

is slow. On the other hand, the exponent of the perpetuity generated by $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, where $(Z_t)_{1\leq t\leq n}$ and $(\epsilon_t)_{1\leq t\leq n}$ are i.i.d., t_{10}-distributed sequences independent of each other, is $\mathbb{E}B_t^2 = 0.01 \times 10/(10-2) = 0.0125$. This leads to a much more ergodic sequence.

10.2.2 Numerical results

Values of the bound calculated treating the parameter point estimates as the true parameters are given in the following tables. The first table deals with the case when $n = 10$, the related threshold u_{10} is the ten-day Value-at-Risk, estimated as described in the previous section, and the value of q_r corresponding to the optimal block size r is also given. The stationary exceedance probability is $p_{10} = \mathbb{P}(X_1 > u_{10}) = 0.001$. Estimates of the various components of the bound are given in Table 10.1, for different values of the m-dependence parameter m and the block parameter r.

Parameter m	2	2	3	3	5
Parameter r	4	8	6	9	10
Bounding term $M_{10}(1)$	105.4	105.4	91.55	91.55	64.38
Bounding term $M_{10}(2)$	0.008	0.016	0.012	0.018	0.02
Bounding term $M_{10}(3)$	0.004357	0.00218	0.00234	0.00234	0.002659
Bounding term $M_{10}(4)$	0.003315	0.004163	0.003739	0.004375	0.004587
Total bound M_{10}	105.5	105.5	91.57	91.57	64.4

Table 10.1: *Values of the bound for different parameters m and r, for the GARCH(1,1) process, with $u_{10} = 129.4$ and $\hat{q}_{10} = 0.001$.*

The bound is very large because of the first term $M_{10}(1)$, corresponding to the m-dependent approximation, because the coefficient of geometric decay $\mathbb{E}B_t^2$ is too close to 1. Let us now have a look at the estimates for the $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$ sequence, gathered in Table 10.2.

Parameter m	2	2	3	3	4	5
Parameter r	4	8	6	9	8	10
$M_{10}(1)$	1.842	1.842	0.3741	0.3741	0.07442	0.01439
$M_{10}(2)$	0.008	0.016	0.012	0.018	0.016	0.02
$M_{10}(3)$	0.007948	0.003978	0.005954	0.005954	0.007928	0.0099
$M_{10}(4)$	0.01053	0.021	0.01579	0.02369	0.02106	0.02619
Bound M_{10}	1.868	1.883	0.4078	0.4217	0.1194	0.07049

Table 10.2: *Values of the bound for different parameters m and r, for the process $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, with $u_{10} = 3.707$ and $\hat{q}_{10} = 0.0102$.*

Because the process is more ergodic, the bound has gone below one, and bears some significance in terms of approximation error, which we shall discuss in the next section. Let us now check that there actually is a convergence to the compound Poisson distribution, and look at sequences of length $n = 10,000$, with exceedance probability $p_{10,000} = 10^{-6}$, and corresponding threshold $u_{10,000}$. The results are presented in Table 10.3.

Parameter m	1,000	1,500	1,800	2,000	2,200	2,200
Parameter r	2,000	3,000	3,600	4,000	4,400	5,000
$M_{10,000}(1)$	48.18	0.9099	0.0856	0.01746	0.00356	0.00356
$M_{10,000}(2)$	0.004	0.006	0.0072	0.008	0.0088	0.01
$M_{10,000}(3)$	0.0975	0.088	0.07072	0.07843	0.0861	0.0861
$M_{10,000}(4)$	0.0532	0.07996	0.09606	0.1068	0.1176	0.1338
Bound $M_{10,000}$	48.33	1.084	0.2596	0.2107	0.216	0.2334

Table 10.3: *Values of the bound for different parameters m and r, for the GARCH(1,1) process, with $u_{10,000} = 155.6$ and $\hat{q}_{4,000} = 0.04$.*

The bound has now also gone below one for the GARCH(1,1) process, which confirms the convergence. Let us now look at the case of the perpetuity generated

by $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$ for $n = 10,000$ and $p_{10,000} = 10^{-6}$, in Table 10.4.

m	10	10	12	13	13	50
r	20	300	280	280	1,000	280
$M_{10,000}(1)$	0.01936	0.01936	0.001041	0.0002414	0.002414	7.889×10^{-28}
$M_{10,000}(2)$	4×10^{-5}	0.0006	0.00056	0.00056	0.002	0.0056
$M_{10,000}(3)$	0.08897	0.005997	0.007755	0.008453	0.002419	0.03408
$M_{10,000}(4)$	0.0005062	0.007931	0.007401	0.007401	0.02653	0.007401
$M_{10,000}$	0.1089	0.03389	0.01676	0.01666	0.03119	0.04203

Table 10.4: *Values of the bound for different parameters m and r, for the process $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, with $u_{10,000} = 7.156$ and $\hat{q}_{280} = 0.0028$.*

Here as well, the bound has shrunk, reaching an approximation error level of 1.7%. In order to illustrate the convergence, there is a trade-off between choosing smaller n to obtain a stronger estimation power, but yet getting a larger bound because of the geometric decay of $M_n(1)$, and choosing a larger n to bring the geometric term $M_n(1)$ down, but yet getting larger, unreliable, values of the transition probability estimates, leading to a larger value of $M_n(2) + M_n(3) + M_n(4)$. In all of these four cases, it is interesting to notice that the fourth term, $M_n(4)$, increases for larger values of r. Also, for a fixed n, there is another trade-off between having a large m to keep $M_n(1)$ small and having a small $r > 2m$, to control $M_n(4)$, derived by Xia [68] with Stein's method, and also $M_n(3)$, derived from a total variation approach.

10.2.3 Approximation error

Let us now discuss the meaning of these numerical results. Taking, for instance, the case of a sequence of length $n = 10$ from the process $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, involved in the calculation of the ten-day VaR, here is what we have:

$$\left|\mathbb{E}\left(\sum_{i=1}^{10}\mathbb{I}_{\{X_i>u_{10}\}}\right) - \mathbb{E}\Big(\mathrm{CP}\big(10p_{10}\theta_{r_{10}}, \nu_{r_{10}}\big)\Big)\right| \leq 10M_{10},$$

and consequently, using the explicit forms of the two expectations from Equation (10.4), we have

$$\begin{aligned}
&\left|10\mathbb{P}(X_1 > u_{10}) - 0.01\right| \leq 10M_{10} \\
&\quad\Longrightarrow\quad 0.001 - M_{10} \leq \mathbb{P}(X_1 > u_{10}) \leq 0.001 + M_{10} \\
&\quad\Longrightarrow\quad 0.001 - M_{10} \leq 1 - \mathbb{P}(X_1 \leq u_{10}) \leq 0.001 + M_{10} \\
&\quad\Longrightarrow\quad -0.999 - M_{10} \leq -\mathbb{P}(X_1 \leq u_{10}) \leq -0.999 + M_{10} \\
&\quad\Longrightarrow\quad 0.999 - M_{10} \leq \mathbb{P}(X_1 \leq u_{10}) \leq 0.999 + M_{10},
\end{aligned}$$

and therefore, noting that the inverse distribution function F_X^{-1} of X is increasing, we have

$$F_X^{-1}\big(0.999 - M_{10}\big) \leq u_{10} \leq F_X^{-1}\big(0.999 + M_{10}\big),$$

which means that $F_X^{-1}\big(0.999 + M_{10}\big) - F_X^{-1}\big(0.999 - M_{10}\big)$ is the *approximation error* made when calculating the ten-day VaR under the assumption that the exceedances form a compound Poisson process. Now, in this case, $0.999 + M_{10} > 1$, and hence the error can be reduced to $F_X^{-1}(1) - F_X^{-1}\big(0.999 - M_{10}\big)$. In the case of the process $X_t = 0.1Z_tX_{t-1} + 0.9\epsilon_t$, and using empirical quantiles from a sequence of length 1,000,000 to estimate the inverse distribution function, with a choice of parameters $m = 5$ and $r = 10$ and the resulting bound estimate of $M_{10} = 0.07049$, *c.f.* Table 10.2.2, we obtain an approximate model error interval of $1.44 \leq u_{10} \leq 7.224$ about the estimated Value-at-Risk of 3.707. In other words, the ten-day Value-at-Risk is comprised between the 92.85% empirical quantile and the empirical maximum. In the case of a sequence of length $n = 10,000$ from the same perpetuity, we get, for the exceedance point process above the 99.9999% quantile, with a choice of parameters $m = 13$ and $r = 280$, a bound for the related compound Poisson approximation of $M_{10,000} = 0.01666$, which implies an approximate model error interval of $2.227 \leq u_{10,000} \leq 7.224$, *c.f.* Table 10.2.2. In other words, the ten-day Value-at-Risk is comprised between the 98.33% empirical quantile and the empirical maximum. Even

though one could hope for more precision, this is still enough to be confident that the relationship $\mathrm{VaR}_{10} = \sqrt{10}\mathrm{VaR}_1$ used in the banking literature is somewhat hazardous. We shall illustrate this point at the end of the next section.

10.3 The approximating compound Poisson process

Let us now focus on the approximating process. From now on, we are not looking anymore at the accuracy of the approximation, but rather at the model itself. In order to estimate compound Poisson parameters properly, *i.e.* the intensity $\theta_r \boldsymbol{\mu}(\bullet)$ and the compounding distribution ν_r, we need longer simulated sequences. This means that we have to lift the constraint $np_n \to t \in \mathbb{R}+$, because we are now in the framework where the threshold u is known, and we want to describe the rate at which the perpetuity exceeds it. For this purpose, we simulate a sequence of length 100,000 from the GARCH(1,1) model, and look at exceedances above the ten-day VaR threshold estimated as the 99.9% quantile $u = 129.4149$, with approximate 99% confidence interval of $[129.3922, 129.4377]$. First, let us look at the corresponding bound, with $n = 100,000$ and $p = 0.001$. As shown on the table below, we need to choose m between 1500 and 2000 to bring the first term $M_n(1) = 8(n-m)\|f_X\|_\infty^{2/3}(\mathbb{E}X^2)^{1/3}(\mathbb{E}B^2)^{m/3}$ of the bound below 1.

Parameter m	15	150	1500	2000	3000
$M_n(1)$	1098000	381400	9.727	0.1935	0.00007658
$M_n(2) + M_n(3) + M_n(4)$	2.070	2.673	8.823	11.11	15.70
Bound M_n	1098000	381400	18.55	11.31	15.70

Table 10.5: *Values of the bound for different parameters m.*

The value of r is chosen to be $2m$. For each estimation, we only retain four significant figures. It is only for m larger than 1500 that $M_n(1)$ becomes of the same order as the rest of the bound. Keeping in mind that the Wasserstein distance is

smaller than 1 anyway, the bound as a whole is not helpful to choose appropriate parameters m and r for smaller sequences. It is just an indication of the convergence speed. Now, there is another way to make such choices.

10.3.1 The parameter m

Recalling that m has been introduced because the perpetuity was approximated by an m-dependent process, we can look at how large m needs to be for us to consider that we are dealing with an m-dependent sequence. This can be achieved by examining the transition probabilities. Indeed, we estimate $\mathbb{P}\big(X_t > u \big| X_{t-1} \leq u, \ldots, X_{t-m} \leq u\big)$ by counting the number of times when a threshold exceedance occurs after m consecutive lags below the threshold, and dividing it by the total number of $(m+1)$-tuples, *i.e.* $n-m$. We calculate this quantity for increasing values of m, and we can realise that they become progressively constant. In other words, no more extra information is gained by taking into account an extra lag. The relevant m to choose is then the value above which changes in these transition probabilities are minor, as is illustrated by the following figure.

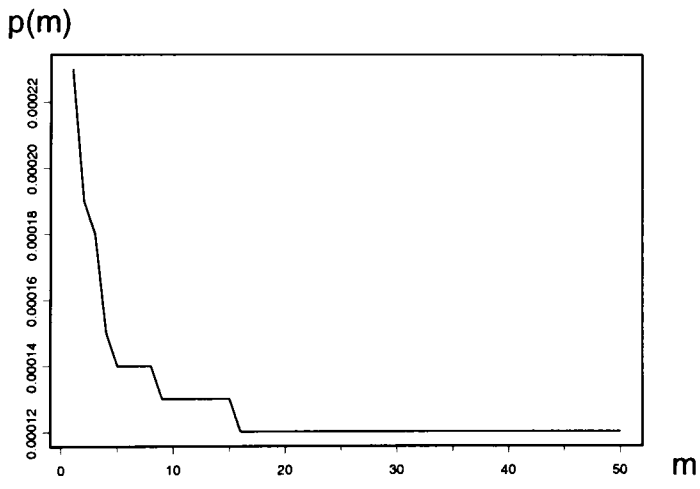

Figure 10.2: *Estimated probability to be above the threshold given that the m previous lags where spent below it, against the parameter m.*

Successive simulations of samples of length 1,000,000 from our perpetuity strongly

suggest that the appropriate m should be larger than 16. We could have guessed anyway that it does not makes much sense to choose huge values for m, even if they reduce the bound. Indeed, there is very little long-range dependence in many financial time series, such as exchange rates, where old data are often useless to describe present behaviours. Usually all relevant information is contained in recent data.

10.3.2 The parameter r

Now, let us focus on the parameter r, *i.e.* the size of the blocks into which we plan to decompose our time series. Recall from the last chapter that, for theoretical reasons, we need to have $r \geq 2m$. Now that we have chosen m, this fixes the first part of the bound, and we can look at the second part, namely $M_n(2) + M_n(3) + M_n(4) = 2rp + 2 - 2\big((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\big)^{\lfloor n/r \rfloor} + q_r\big(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r}\big)$. This is minimised by taking r as small as possible, just because the term $2rp$ is the biggest for non-trivial values of r. This is due to the fact that p is fixed, because n and u are. This might indicate $r = 2m$ as an appropriate choice. In order to check whether this is actually the case, let us have a closer look at our time series. Recall that the compounding distribution of our final compound Poisson model is given by $\mathcal{L}\big(T_{r,1} \big| T_{r,1} \geq 1\big)$. In other words, this is the number of exceedances per block of size r, whenever they actually occur. As a result, empirically, a good guess for r would be the size of the biggest exceedance clump. So if this is not too far from $2m$, this means that we are making the right choice. And indeed, simulating several GARCH samples of length 1,000,000, for most of them the biggest cluster of exceedances is no larger than 60. The following plot shows the distribution of exceedance clump sizes in a simulated sample.

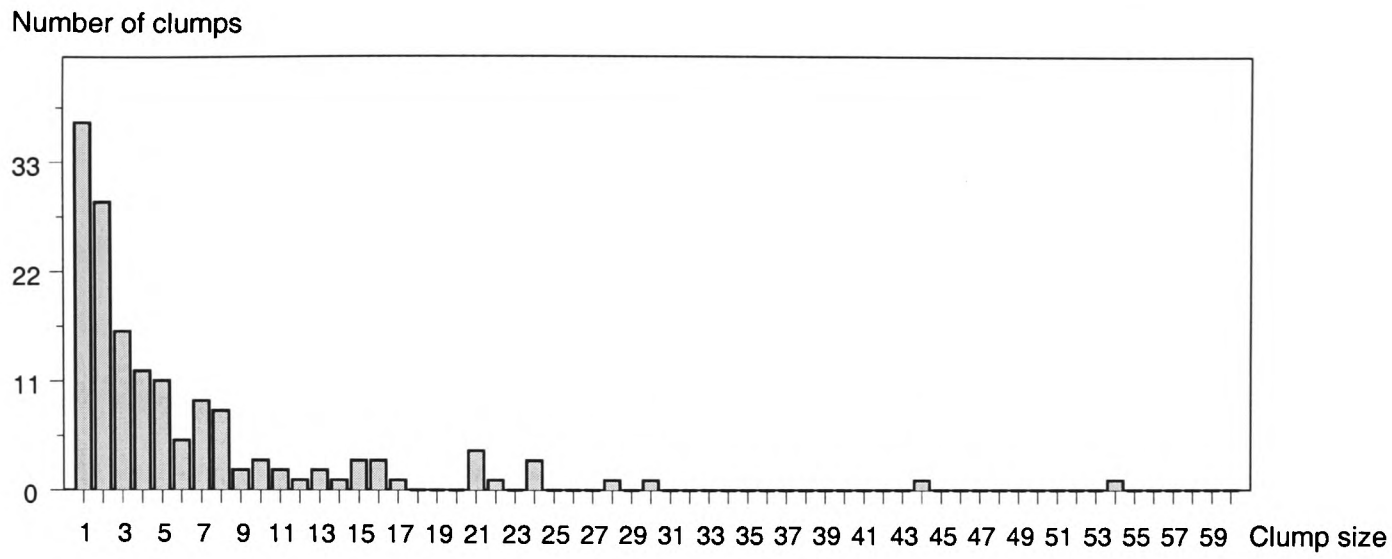

Figure 10.3: *Number of exceedance clumps of size r, against r, recorded in a simulated GARCH sample of length 1,000,000.*

10.3.3 Simulation of the compound Poisson process

The aim is to simulate the compound Poisson process describing exceedances above the VaR threshold, in order to check its accuracy. We are looking at $\mathrm{CP}(10p\theta_r, \nu_r)$. In this practical setting, since we select the parameter m so that the process simulated be very close to m-dependence, we approximate q_r and the transition probabilities p_k associated with the process $(\widetilde{X}_t)$, as defined in Theorem 7.4.1, by those given by the process (X_t). The estimator for q_r is given by formula (10.5). Given that $\theta_r = \frac{\mathbb{P}(T_{r,1}\geq 1)}{rp}$, the intensity is $\frac{10\mathbb{P}(T_{r,1}\geq 1)}{r}$. Now the same comments also apply to the compounding distribution. We have an exact expression for it, but it depends on parameters whose estimates are rather unreliable, because they are calculated from scarce data. Hence, it is wiser to evaluate the end product directly. For this purpose, we shall build a pool of values in order to constitute an empirical clump distribution. Yet, one simulated GARCH sample of length 1,000,000 will only give around 15,000 blocks, amongst which usually only about 800 contain exceedances. Nevertheless, by simulating more samples, this empirical collection of possible clump sizes can be enlarged at will. We shall iterate the operation until the pool reaches the size of about 20,000. This gives us an empirical cluster size distribution.

10.3.4 The compound Poisson estimates

Finally, we simply have to pick a random sample from the pool mentioned above, whose length we set to be a Poisson($\frac{10\hat{q}_r}{r}$) random variable, and sum the elements thus obtained. This gives us our approximation to the number of exceedances over the VaR threshold in ten days. Now simulating a sequence of length 1,000,000 from this distribution, and an empirical compounding distribution of 24149 points, we get the estimate $\hat{\mu} = 0.010128$, with an approximate 99% confidence interval of $[0.008962177, 0.01129382]$. Given that the mean was supposed to be 0.01, this confirms the validity of our model.

10.4 Multivariate case

Multidimensional perpetuities are powerful tools to describe joint evolutions, but there are a few things they cannot do. Despite what would seem logical, they do not encompass multivariate GARCH models. Indeed, even though the conditional variance of a one-dimensional GARCH(1,1) model does form a perpetuity, the conditional covariance matrix of a multivariate GARCH model does not follow a stochastic recursion as we defined them. Indeed, a multivariate GARCH model describes the evolution of the covariance matrix of a vector. Now, the theory developed in this thesis is for linear recursions on vectors, not on matrices. This is why, for the purpose of the illustration of the multidimensional case, we shall have to choose an example. We shall look at

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} B_{1,1,t} & B_{1,2,t} \\ B_{2,1,t} & B_{2,2,t} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} A_{1,t} \\ A_{2,t} \end{pmatrix}, \tag{10.5}$$

where $B_{1,1,t} = 0.4Z^2_{1,1,t}$, $B_{1,2,t} = 0.2Z^2_{1,2,t}$, $B_{2,1,t} = 0.4Z^2_{2,1,t}$, $B_{2,2,t} = 0.3Z^2_{2,2,t}$, $A_{1,t} = 0.4Z^2_{1,t}$ and $A_{2,t} = 0.2Z^2_{2,t}$. The random variables $Z_{1,1,t}$, $Z_{1,2,t}$, $Z_{2,1,t}$, $Z_{2,2,t}$, $Z_{1,t}$ and $Z_{2,t}$

are chosen to be i.i.d. t_{10}. Such a recursion verifies the strict stationarity conditions of Kesten. Here is a plot of a simulated sample generated by it.

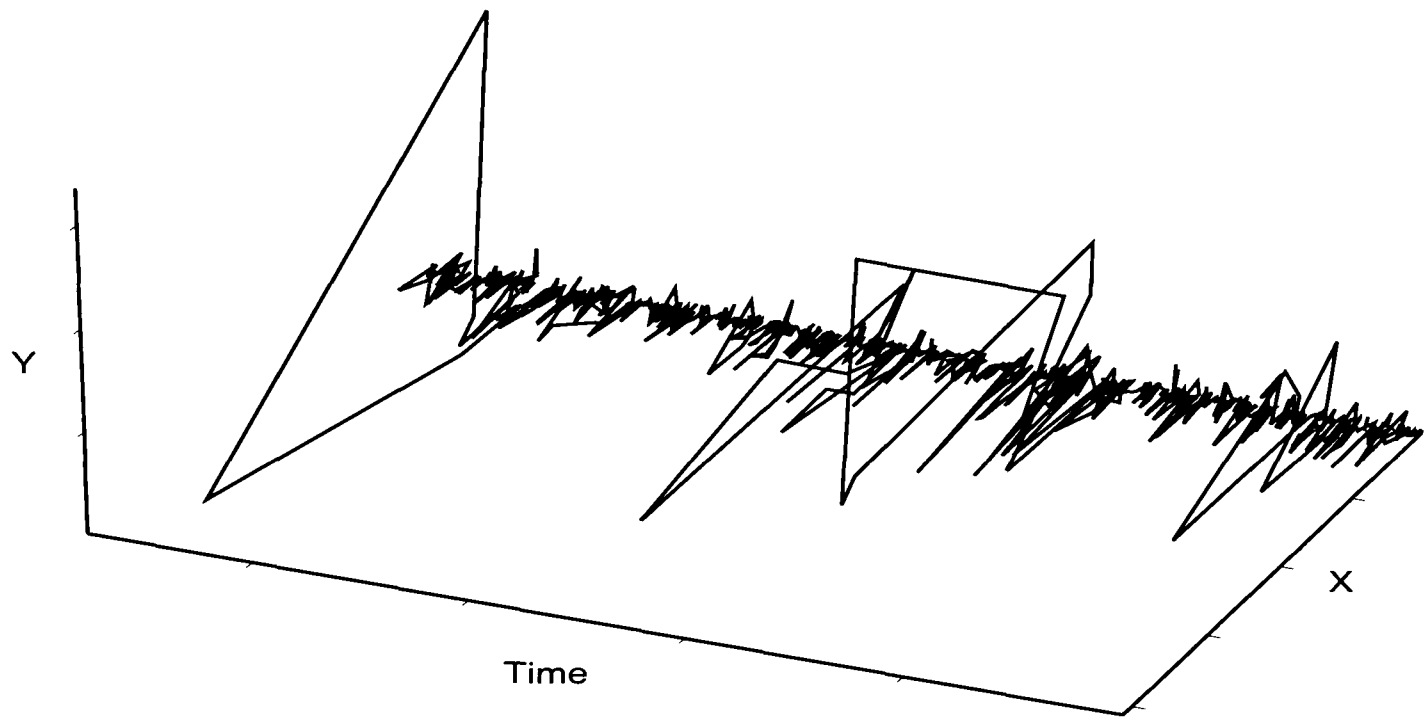

Figure 10.4: *View of the perpetuity from above. The width is X_t, the height is Y_t, and time increases from right to left.*

10.4.1 Bivariate exceedances

Now, we are interested in exceedances of the vector (X_t, Y_t) beyond a rectangular border (u_X, u_Y). An obvious question is the choice of thresholds we might want to look at. In this multidimensional setting, VaR considerations would make us look at, say, a portfolio composed of assets X and Y, in given fixed quantities. The portfolio VaR is then the overall portfolio value (or loss, or return) that has 1% chances to be exceeded within ten days. A serious obstacle for us is that, even though our covariates follow a multidimensional stochastic recursion, the overall portfolio, *i.e.* the weighted sum of the two securities, does not necessarily follow a one-dimensional stochastic recursion as a result. Hence this present work will only tell us how the covariates exceed a given multidimensional threshold, but nothing precise about their weighted

sum. Indeed, let us take the simple case of $u_X = 15$ and $u_Y = 20$, and a portfolio composed of three securities of type X and three of type Y. If (X_t, Y_t) exceeds $(15, 20)$, this means that the portfolio exceeds the threshold $3 \times 15 + 3 \times 20 = 105$. But this would also have happened for $X = 5$ and $Y = 30$, which corresponds to another multivariate border. This illustrates how portfolio exceedances do not relate easily to multivariate exceedances. Indeed, if we denote by (w_x, w_y) the row vector of weights, then the value V_t of the portfolio at time t is given by $V_t = (w_X, w_Y)^T(X_t, Y_t)$. Now, if (X_t, Y_t) follow a bidimensional stochastic recursion of the type (10.5), then we have

$$V_t = \begin{pmatrix} w_X & w_Y \end{pmatrix} \begin{pmatrix} B_{1,1,t} & B_{1,2,t} \\ B_{2,1,t} & B_{2,2,t} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} w_X & w_Y \end{pmatrix} \begin{pmatrix} A_{1,t} \\ A_{2,t} \end{pmatrix},$$

and furthermore,

$$\begin{aligned} &\begin{pmatrix} w_X & w_Y \end{pmatrix} \begin{pmatrix} B_{1,1,t} & B_{1,2,t} \\ B_{2,1,t} & B_{2,2,t} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} w_X B_{1,1,t} + w_Y B_{2,1,t} & w_X B_{1,2,t} + w_Y B_{2,2,t} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \\ &= \big(w_X B_{1,1,t} + w_Y B_{2,1,t}\big) X_{t-1} + \big(w_X B_{1,2,t} + w_Y B_{2,2,t}\big) Y_{t-1}. \end{aligned}$$

Now, this is only equal to $V_{t-1} = w_X X_{t-1} + w_Y Y_{t-1}$ if $B_{1,2,t} = B_{2,1,t} = 0$ almost surely and $B_{1,1,t} = B_{2,2,t}$ almost surely, *i.e.* if $\boldsymbol{B}_t$ is a homothety. In the general case when there is no constraint on the shape of $\boldsymbol{B}_t$, we are still in a position to describe the rate at which a multidimensional perpetuity exceeds a multivariate border. The procedure is quite similar to the one of the one-dimensional case. We simply have to simulate the bivariate perpetuity, and record whenever it exceeds the two dimensional threshold chosen.

10.4.2 Evaluation of the bound

The theoretical bound is

$$
\begin{aligned}
& d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{\mathbf{0},\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right) \\
&\leq \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2+2(n-m)\left(\prod_{i=1}^{d}(u_i+\epsilon_m)-\prod_{i=1}^{d}u_i\right)\left\|f_{\boldsymbol{X}}\right\|_\infty \\
&\quad +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right).
\end{aligned}
$$

Here, $d=2$, hence we get

$$
\begin{aligned}
& d_W\left(\mathcal{L}\left(\sum_{t=1}^{n}\mathbb{I}_{\{\boldsymbol{X}_t\notin R_{\mathbf{0},\boldsymbol{u}}\}}\delta_{tp}(\bullet)\right),\mathrm{CP}\left(\theta_r\boldsymbol{\mu}(\bullet),\nu_r\right)\right) \\
&\leq \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2+2(n-m)\epsilon_m(\epsilon_m+u_X+u_Y)\left\|f_{\boldsymbol{X}}\right\|_\infty \\
&\quad +2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right).
\end{aligned}
$$

As mentioned after the proof of Theorem 9.2.2, we denote

$$
\begin{aligned}
\mathcal{M}_n(1) &= \frac{8(n-m)}{\epsilon_m^2}\left(\mathbb{E}\|\boldsymbol{B}\|^2\right)^m\mathbb{E}\|\boldsymbol{X}\|_2^2 \\
\mathcal{M}_n(2) &= 2(n-m)\epsilon_m(\epsilon_m+u_X+u_Y)\left\|f_{\boldsymbol{X}}\right\|_\infty \\
\mathcal{M}_n(3) &= 2rp \\
\mathcal{M}_n(4) &= 2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor\frac{n}{r}\rfloor} \\
\mathcal{M}_n(5) &= q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right),
\end{aligned}
$$

and $\mathcal{M}_n=\mathcal{M}_n(1)+\mathcal{M}_n(2)+\mathcal{M}_n(3)+\mathcal{M}_n(4)+\mathcal{M}_n(5)$. We can find an optimal value for ϵ_m numerically, as described in Remark 2 following Theorem 8.5.1. Now, we need to find a numerical solution to this. $\mathbb{E}\|\boldsymbol{X}\|_2^2$ is estimated from the sample, $\left\|f_{\boldsymbol{X}}\right\|_\infty$ is roughly estimated as the maximum of the kernel density. Now, as we are in a two-dimensional framework, the kernel density estimator used is

$$
\hat{f}(x,y)=\frac{\sum_s\Phi\big((x-x_s)/h_x\big)\Phi\big((y-y_s)/h_y\big)}{nh_xh_y},
$$

where Φ is the standard normal distribution function, (x_s, y_s) the points at which the density is estimated, h_x and h_y the bin widths from the average shifted histogram method, *c.f.* Venables and Ripley [65], Section 5.6. Given that $|||\boldsymbol{B}|||$ is the largest eigenvalue of $\boldsymbol{B}$ with respect to the Euclidian norm, we can run 100,000 simulations of it and compute the related estimate of $\mathbb{E}|||\boldsymbol{B}|||^2$, obtaining a figure of 0.1901974, with approximate 99% confidence interval of $[0.1894465, 0.1909483]$. Let us look at the convergence to the compound Poisson model, as in Section 10.2. For this purpose, we shall calculate the bound for a sequence of length $n = 10$, and for one of length $n = 100$, with appropriate exceedance probabilities in order to satisfy the condition $n\mathbb{P}(\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}_n}) \to t \in]0, +\infty[$ exposed in the Remark following Theorem 9.2.2. We shall choose our two-dimensional thresholds such that $n\mathbb{P}(\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}_n})$ stays of the same order. The results are gathered in the following table, where the bound is computed for the optimal parameters m, ϵ and r that minimize it numerically. We denote $p_n = \mathbb{P}(\boldsymbol{X}_t \notin R_{0,\boldsymbol{u}_n})$.

n	u_X	u_Y	p_n	$\mathcal{M}_n(1)$	$\mathcal{M}_n(2)$	$\mathcal{M}_n(3)$	$\mathcal{M}_n(4)$	$\mathcal{M}_n(5)$	$\mathcal{M}_n$
10	70	80	0.00185	3600	6910	0.056	0.0192	0.0467	10500
100	180	300	0.00028	5.4×10^{-7}	1.08×10^{-6}	0.056	0.0155	0.0407	0.112

Table 10.6: *Values of the bound for different parameters.*

This clearly illustrates the validity of the approximation in the multidimensional case. Already for a sequence of length 100, which is not very long, we only get an approximation error of 11.2%.

Now that we can be satisfied with the approximation, let us focus on longer sequences, in order to be in a position to describe the compound Poisson behaviour of the exceedances properly, lift the constraint that np_n should be more or less constant, and estimate the model parameters, as in Section 10.3. Let us choose a two-dimensional threshold of $(u_X, u_Y) = (70, 80)$. It has an estimated probability of about 0.00185% of being exceeded. Let us fix this exceedance probability and look

at the related compound Poisson model. Firstly, we can calculate the bound for the sequence of length $n = 100,000$ that we shall use later on. For each value of m, the optimal ϵ, and then the optimal r are computed.

Parameter m	10	42	45	50	100	100
Parameter ϵ	0.000769	10^{-11}	10^{-12}	10^{-13}	10^{-25}	10^{-25}
Parameter r	100	84	90	290	200	20000
$\mathcal{M}_{100,000}(1)$	6200000	0.341	0.271	0.00613	6.1×10^{-15}	6.1×10^{-15}
$\mathcal{M}_{100,000}(2)$	19100000	0.308	0.046	0.00341	2.8×10^{-15}	2.8×10^{-15}
$\mathcal{M}_{100,000}(3)$	0.406	0.341	0.365	0.406	0.812	81.2
$\mathcal{M}_{100,000}(4)$	2	2	2	2	2	0.755
$\mathcal{M}_{100,000}(5)$	0.258	0.215	0.23	0.256	0.514	18200
Bound $\mathcal{M}_{100,000}$	25300000	3.2	2.91	2.67	3.33	18300

Table 10.7: *Values of the bound for different parameters.*

The term $\mathcal{M}_{100,000}(3) + \mathcal{M}_{100,000}(4) + \mathcal{M}_{100,000}(5)$ of the bound is the one that prevents it from easily falling below one. This is due to the fact that it needs a very low probability of exceedance to do so, whereas the estimates we are using are calculated from simulated data, which limits how small they can be. Indeed, with a finite sequence, fixing too high a threshold would return the estimate 0. Again, we are faced with the trade-off between choosing a large m for the mixing part of the bound to be small, and choosing r not too large to control the Bernstein blocks part of the bound.

10.4.3 Simulation of the compound Poisson process

As in the one-dimensional case, we are faced with the same contingencies concerning the robustness of the transition probability estimators, and the choices of r and m. The same procedure shall be adopted, this time using the sequence of border

exceedances. More precisely, we are recording the times at which the process (X_t, Y_t) is outside the rectangle $R_{0,\boldsymbol{u}}$, *i.e.* whenever either $X_t \geq u_X$, or $Y_t \geq u_Y$. In this practical setting, since we select the parameter m so that the process simulated be very close to m-dependence, we approximate q_r and the transition probabilities $p_{\boldsymbol{k}}$ associated with the process $\big((\widetilde{X}_t, \widetilde{Y}_t)\big)$, as defined in Theorem 8.5.1, by those given by the process $\big((X_t, Y_t)\big)$. Choosing, as an example, $u_X = 100$ and $u_Y = 150$, we get an average number of exceedances of 0.627, with approximate 99% confidence interval $[0.601, 0.653]$.

10.4.4 Portfolio VaR

One possible way to evaluate the VaR of a portfolio with this method is to consider that each of the two covariates forms a one-dimensional perpetuity, and that they share the same multiplicative factor B_t in their random recursion. In other words, this boils down to considering a stochastic recurrence equation with homothetic $\boldsymbol{B}_t$. In Equation (10.5), let us keep the same random coefficients, except for $B_{1,2,t}$ and $B_{2,1,t}$ that we put equal to zero, and $B_{1,1,t} = B_{2,2,t} = 0.4Z_t^2$, where (Z_t) is a sequence of i.i.d., t_{10} random variables. In such a new situation, the value of a portfolio of assets X and Y, *i.e.* a linear combination of them, follows exactly the same random recurrence equation. Hence the whole procedure exposed previously applies to it. Let us, for instance, look at a portfolio of $w_X = 30$ shares of security X and $w_Y = 50$ shares of security Y. We can now run the whole one-dimensional procedure for $w_X X_t + w_Y Y_t$. Hence we obtain for the 99%-confidence, ten-day VaR an estimate of VaR $= 213.5$ and simulating from a sequence of length 100,000 from its empirical compound Poisson distribution, we get the estimate $\hat{\mu} = 0.01$, with an approximate 99% confidence interval of $[0.0093, 0.012]$, which confirms the validity of the model, in so far as this is the average number of exceedances above the VaR threshold.

10.4.5 Quantitative considerations

So far, the cases we examined led us to values of the bound greater than one. This has prevented us from using it as a way to assess the accuracy of the compound Poisson approximation, simply because a Wasserstein distance between probability distributions must be smaller than one by definition. As shown by Tables 10.3 and 10.4.2, there are two reasons why the bound might be too large. First, the process might not be ergodic enough, having a coefficient $\mathbb{E}|||\boldsymbol{B}|||^2$ of geometric decay too close to 1. This would make the process mix too slowly, giving large values for the first term of the bound $8(n-m)\epsilon_m^{-2}\left(\mathbb{E}|||\boldsymbol{B}|||^2\right)^m \mathbb{E}\|\boldsymbol{X}\|_2^2$. Second, the threshold might not be high enough, and hence give block exceedance and transition probabilities, which are too large, and we would then obtain a large value for the second term of the bound $2(n-m)\epsilon_m(\epsilon_m+u_X+u_Y)\left\|f_{\boldsymbol{X}}\right\|_\infty+2rp+2-2\left((1-p)\prod_{i=1}^{m-1}(1-p_{\boldsymbol{v}_i})\right)^{\lfloor n/r\rfloor}+q_r\left(\frac{1.65}{\sqrt{1-q_r}}+e^{q_r}\right)$. There are ways to remedy to this. Concerning the first point, we can look at processes which are more ergodic, by decreasing the multiplicative coefficient $\boldsymbol{B}$. In the one-dimensional example mentioned before, this can be achieved by looking at $B_t = 0.1Z_t$, where Z_t is a t_{10} random variable, instead of $B_t = 0.8Z_t$, say. About the second point, it is useful to note here that we have so far been dealing with daily data. Now, if the time points correspond to intra-day data, say, for instance, ten points per day, then a period of ten days corresponds to one hundred time points. As a result, according to Equation (10.4), the relevant ten-day, 99%-confidence VaR threshold is the 99.99% quantile, and not the 99.9% one. This raises the threshold considered, and consequently reduces the values of the probabilities used in the second part of the bound. With these comments in mind, let us look back on our bound estimates, for the compound Poisson approximation of the point process of exceedances above the ten-day, 99%-confidence VaR threshold, with a frequency of ten time points per day. They are calculated for different perpetuities, of which the 99.99% quantile is taken as threshold. The values of the parameters m and r displayed, corresponding to m-dependence and block size, are those giving the smallest, and hence best, value

of the bound for the process of interest. A few examples are gathered in the following table. Here, $\left(Z_t^{(d)}\right)$ and $\left(\varepsilon_t^{(d)}\right)$ denote two i.i.d. series of Student random variables with d degrees of freedom, independent of each other.

$X_{t+1} = 0.1Z_t^{(10)}X_t + 0.7\varepsilon_t^{(10)}$	Bound$(m=11, r=630) = 0.66$
$X_{t+1} = 0.1Z_t^{(10)}X_t + 0.9\varepsilon_t^{(10)}$	Bound$(m=11, r=630) = 0.66$
$X_{t+1} = 0.1Z_t^{(4)}X_t + 0.9\varepsilon_t^{(4)}$	Bound$(m=12, r=630) = 0.70$
$X_{t+1} = 0.01Z_t^{(10)}X_t + 0.9\varepsilon_t^{(10)}$	Bound$(m=6, r=470) = 0.47$

Table 10.8: *Values of the bound for various perpetuities.*

The bound has gone below one despite the length ($n = 100,000$) of the sequences. Now, this means that for the related perpetuities, the distribution of the exceedance point process and the one of the approximating compound Poisson measure are Wasserstein-close to one another. Hence the approximation makes sense, which was already clear from the results of Section 10.2.2.

This gives precious insight into the time evolution of these exceedance processes. In particular, we can step away from the classic approach in the banking literature, *c.f.* Crouhy *et al.* [22], were the ten-day VaR is derived as $\sqrt{10}$ times the one-day VaR, after assuming that:

- The returns are Gaussian, and therefore the VaR is proportional to their standard deviation
- The returns are i.i.d.

Both these hypotheses are notoriously hazardous, and it is therefore a relief to be able to lift them away. Besides, if we consider the last perpetuity of Table 10.4.5, with the present compound Poisson approach, for daily time points, the one-day VaR is the 99% quantile, *i.e.* $u = 2.48$, and the ten-day VaR is the 99.9% quantile, *i.e.* $u = 3.72$. These quantiles are estimated from the tail of the stationary distribution, which is regularly varying according to Kesten's theorem. Now, 3.72 is smaller than

$\sqrt{10} \times 2.48 = 7.84$. This means that banks tend to exaggerate the spread between one-day VaR and ten-day VaR, and hence risk to overestimate their ten-day VaR. Indeed, a bank would here consider that 1% of its ten-day losses exceed 7.84, whereas the compound Poisson method tells them that it is only 3.72 that is exceeded with probability 1%. Now, risk underestimation is obviously perillous, but overestimating ones risk is also dangerous, because then banks become exceedingly risk-averse. Consequently, whenever the value of a risky asset starts to fall, there will be no-one to acquire it, and its downfall will continue, and even accelerate. Such an argument is exposed in Danìelsson *et al.* [23], yet not exactly from the same angle, given that the authors rather use it to illustrate the danger of homogenizing the risk-aversion of agents. This is why too large a risk measure will threaten market stability. Yet, these comments must be balanced by the observation from Borkovec and Klüppelberg [17] that banks usually underestimate their one-day VaRs, because of their Gaussian assumptions. So it might happen that the underestimation of one-day VaR, coupled with an overestimated spread, give a reasonable ten-day VaR figure. But then this would be pure luck.

Chapter 11

Conclusion

This thesis provides a compound Poisson approximation for the occurrence of threshold exceedances in the course of a finite multidimensional stochastic recursion. We are indeed dealing with finite sequences. The advantage of such an approach, distinct from any asymptotic consideration, is that the parameters of the approximating model are easier to compute, and that these models can be optimised according to the size of the sequence they describe.

The approximation is built in two steps. Firstly, the recursive nature of the perpetuity is employed to link its exceedance point process to a $\{0,1\}$-valued, m-dependent process. Then the latter is approximated by a compound Poisson random measure, whose parameters are explicitly given after an adaptation to d-th order Markov chains of results by Barbour, Novak and Xia [7] and Goodman [38]. Barbour *et al.* [7] use the method of Bernstein blocks, popular in the Russian school, *c.f.* Ibragimov and Linnik [42], to derive a compound Poisson approximation. In this

thesis, the blocks method is first adapted to the easier case of $\{0,1\}$-valued, first order Markov chains, then to d-th order Markov chains, which encompass d-dependent processes, and therefore enable us to put the final touch to the whole approximation of perpetuity exceedances.

We then obtain a compound Poisson measure, whose intensity and compounding distribution are functions of two free parameters r and m. Both are introduced by the blocks approach, yet the choice of m relies heavily on the properties of the stochastic recursion that generates the perpetuity. The idea is to find the values r^* and m^* of the block size r and the jump lag m for which the bound is minimal, in a given configuration. This can be interpreted as a model selection procedure. First, r^* and m^* have to be found, either numerically as is the case here, or theoretically, by differentiation. Yet, this would require the exact values of the probabilities p and p_k, which are not known as of now. This is why they are estimated. Then, r^* and m^* give us an "optimal", explicit, empirical model $\mathrm{CP}\big(\theta_{r^*}\boldsymbol{\mu}(\bullet), \nu_{r^*}\big)$. The method has the advantage of providing analysts with a model for exceedances, which is especially suited for the finite sequence they are dealing with, with easily computed parameters.

In order to be able to *calculate*, and not just estimate, the parameters of the model, namely θ_{r^*} and ν_{r^*}, we have to know the exact values of p and p_k, in other words the exact form of the distribution of a perpetuity. The estimation of such distributions is studied in Aebi *et al.* [1], where the authors use bootstrap techniques with the "contraction method" for aymptotics of recursive algorithms presented in Rachev and Rüschendorf [59]. The rate of convergence of a discounted sum to its limit distribution has been investigated in Rachev and Samorodnitsky [60], and Brandt [19] has looked into the stability of the solution. As to the distribution itself, results can be found in Dufresne [28] for the continuous-time case, and the main updates for the discrete case are in Goldie and Grübel [37], where perpetuities with thin tails, exponential or Poissonian, are investigated. This is a continuation of Vervaat's work [67], where properties of a perpetuity are derived from conditions on the random

coefficients generating it. The exact distribution of such discounted sums is accessible when coefficients generating the recursion have a discrete distribution, but only the asymptotic behaviour is yet known when the latter is continuous. The present method could therefore be rigorously applied in the case of discrete coefficients, where p and p_k could be exactly calculated, leading to exact values of r^* and m^*, which in turn would give explicit expressions of θ_{r^*} and ν_{r^*}.

Even though this approach is developed for perpetuities taking values on the whole real line in the one-dimensional case, for technical reasons it is restricted to non-negative vectors in the multidimensional setting. The threshold exceedances considered here are excursions outside a d-dimensional parallelepiped of the form $[0, u_1] \times \cdots \times [0, u_d]$ in $\mathbb{R}_d^+$. We describe the pace at which the perpetuity exceeds this zone. Now this method does not account for cases when just a few covariates are large. It would be interesting to try and generalise the proof of Theorem 8.5.1 to the case when the state space is the whole $\mathbb{R}^d$. Another type of exceedance would then have to be considered. Recent developments in the asymptotic multidimensional case can be found in Klüppelberg and Pergamenchtchikov [46], for the special case of AR(q) processes, where the positivity condition on the stochastic coefficients is lifted.

From the viewpoint of extreme value theory, there are two aspects under which multidimensional perpetuities can be observed. Firstly, their recursive nature supplies the covariates with a specific joint dependence structure, and their stationary joint distribution can be described using the concept of regular variation, as in Basrak *et al.* [12]. The second aspect is their time dependence. This is where the compound Poisson approach and this thesis come in.

On the practical side, our compound Poisson approximation enables us to equate directly the ten-day VaR of a sequence of daily data to the 99.9% quantile of their stationary distribution, and the one-day VaR to the 99% quantile, which yields a much narrower spread between the two than in the i.i.d. Gaussian approach, popular in the banking world. In that case, the ten-day VaR is indeed derived as $\sqrt{10}$ times

the one-day VaR. Yet, this potential overestimation of the ten-day VaR is conjugated with a generic underestimation of quantile returns, as illustrated in Borkovec and Klüppelberg [17]. This double mistake is bound to generate untidy outcomes. An advantage of the present compound Poisson approach is that it allows us to identify the VaR for time horizons that differ from the data frequency. Once the relevant quantile to use has been pointed out with this method, it can then be estimated with Formula (28) of Borkovec and Klüppelberg [17].

There are two main other tracks that could be explored further. Firstly, in Section 8.3 concerning the type of recurrence satisfied by each covariate of a perpetuity, some algebraic work could be attempted to find out an explicit expression of the recursive coefficients in each marginal d-th order random recursion, as a function of the original $\big((\boldsymbol{A}_t, \boldsymbol{B}_t)\big)_{1\leq t\leq n}$. This could be even more revealing about how the introduction of additional covariates modifies the marginal behaviours of existing ones. For instance, their original mixing properties could be precisely related to their new ones. An important other track of research would be the further exploration of the exceedance behaviour of a portfolio, say a linear combination of random variables representing assets that form a multidimensional perpetuity. The resulting portfolio is a one-dimensional perpetuity in the homothetic case, but it would be interesting to learn about its exceedance point process in the case of a general random recursion.

Bibliography

[1] AEBI, M., EMBRECHTS, P. AND MIKOSCH, T. (1994) Stochastic discounting, aggregate claims, and the bootstrap. *Adv. Appl. Probab.* **26**, 183-206.

[2] ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1990) Poisson approximation and the Chen-Stein method. *Statist. Science* **5**, 403-434.

[3] ATHREYA, K.B. AND PANTULA, S.G. (1986) Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23**, 880-892.

[4] BARBOUR, A.D., CHEN, L.H.Y. AND LOH, W.-L. (1992) Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20**, 1843-1866.

[5] BARBOUR, C. AND CHRYSSAPHINOU, O. (2001) Compound Poisson approximation: a user's guide. *Ann. Appl. Probab.* **11**, 964-1002.

[6] BARBOUR, A.D., HOLST, L. AND JANSON, S. (1992) *Poisson Approximation.* Oxford University Press, New York.

[7] BARBOUR, A.D., NOVAK, S.Y. AND XIA, A. (2002) Compound Poisson approximation for the distribution of extremes. *Adv. Appl. Probab.* **34**, 223-240.

[8] BARBOUR, A.D. AND UTEV, S. (1999) Compound Poisson approximation in total variation. *Stoch. Proc. Appl.* **82**, 89-125.

[9] BASEL COMMITTEE ON BANKING SUPERVISION (1988) *International Convergence of Capital Measurement and Capital Standards.* Basel Committee Publications **4**, Basle.

[10] BASEL COMMITTEE ON BANKING SUPERVISION (1996) *Amendment to the Capital Accord to Incorporate Market Risks.* Basel Committee Publications **24**, Basle.

[11] BASEL COMMITTEE ON BANKING SUPERVISION (2004) *International Convergence of Capital Measurement and Capital Standards: a Revised Framework.* Basel Committee Publications **107**, Basle.

[12] BASRAK, B., DAVIS, R.A. AND MIKOSCH, T. (2002) A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12**, 908-920.

[13] BENJAMIN, N.A. (2001) *Analysis of Exchange Rate Time Series, with Special Emphasis on Value-at-Risk.* M.Sc. Dissertation. University of Oxford.

[14] BERBEE, H.C.P. (1979) *Random Walks with Stationary Increments and Renewal Theory.* Mathematical Centre Tracts **112**. Amsterdam.

[15] BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987) *Regular variation.* Cambridge University Press, New York.

[16] BOLLERSLEV, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307-327.

[17] BORKOVEC, M. AND KLÜPPELBERG, C. (2000) Extremwerttheorie für Finanzzeitreihen - ein unverzichtbares Werkzeug im Risikomanagement. In: Rudolph, B. and Johanning, L. (Hrsg.) *Handbuch Risikomanagement*, pp. 219-241. Uhlenbruch Verlag, Bad Soden.

[18] BOWMAN, A. AND AZZALINI, A. (1997) *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford University Press, Oxford.

[19] BRANDT, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* **18**, 211-220.

[20] CAPÉRAÀ, P., FOUGÈRES, A.-L. AND GENEST, C. (1997) Estimation of bivariate extreme value copulas. *Biometrika* **84**, 567-577.

[21] CLEVELAND, R.B., CLEVELAND, W.S., MCRAE, J.E. AND TERPENNING, I. (1990) STL: a seasonal-trend decomposition procedure based on loess. *J. Official Statist.* **6**, 3-73.

[22] CROUHY, M., GALAI, D. AND MARK, R. (2001) *Risk Management.* McGraw-Hill, New York.

[23] DANÍELSSON, J., EMBRECHTS, P., GOODHART, C., KEATING, C., MUENNICH, F., RENAULT, O. AND SHIN, H.S. (2001) *An academic response to Basel II.* Financial Markets Group, Special Paper **130**. London School of Economics.

[24] DAVIS, R.A. AND HSING, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879-917.

[25] DAVIS, R.A. AND MIKOSCH, T. (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049-2080.

[26] DAVYDOV, Y.A. (1973) Mixing conditions for Markov chains. *Theory Probab. Appl.* **18**, 312-328.

[27] DOUKHAN, P. (1994) *Mixing: Properties and Examples.* Lecture Notes in Statistics, 85. Springer, New York.

[28] DUFRESNE, D. (1990) The distribution of a perpetuity, with applications to risk theory and pension funding. *Scand. Actuar. J.* **1-2**, 39-79.

[29] EMBRECHTS, P. AND GOLDIE, C.M. (1994) Perpetuities and random equations. In: Mandl, P. and Huskova, M. (Eds) *Asymptotic Statistics. Proceedings of the 5th Prague Symposium, September 4-9, 1993*, pp. 75-86. Physica-Verlag, Heidelberg.

[30] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[31] EMBRECHTS, P., LINDSKOG, F. AND MCNEIL, A. (2003) Modelling dependence with copulas and applications to risk management. In: Rachev, S.T. (Ed) *Handbook of Heavy Tailed Distributions in Finance*, pp. 329-384. Elsevier North-Holland, New York.

[32] ENGLE, R.F. (1982) Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

[33] FEIGIN, P.D. AND TWEEDIE, R.L. (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6**, 1-14.

[34] FELLER, W. (1971) *An Introduction to Probability Theory and Its Applications*, vol. I. Wiley, New York.

[35] GINÉ, E., GÖTZE, F. AND MASON, D.M. (1997) When is the Student t statistic asymptotically standard normal? *Ann. Probab.* **25**, 1514-1531.

[36] GOLDIE, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126-166.

[37] GOLDIE, C.M. AND GRÜBEL, R. (1996) Perpetuities with thin tails. *Adv. Appl. Probab.* **28**, 463-480.

[38] GOODMAN, L.A. (1958) Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains. *Ann. Math. Stat.* **29**, 476-490.

[39] HAAN, L. DE (1985) Extremes in higher dimensions: The model and some statistics. In: *Bulletin of the International Statistical Institute*, Amsterdam. Proceedings of the 45th Session of the International Statistical Institute.

[40] HEFFERNAN, J.E. AND TAWN, J.A. (2004) A conditional approach for multivariate extreme values. *J. Roy. Statist. Soc. Ser. B* **66**, 497-546.

[41] HSING, T., HÜSLER, J. AND LEADBETTER, M.R. (1988) On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78**, 97-112.

[42] IBRAGIMOV, I.A. AND LINNIK, Y.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[43] KALLENBERG, O. (1997) *Foundations of Modern Probability.* Springer, New York.

[44] KESTEN, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207-248.

[45] KLÜPPELBERG, C. (2001) Developments in insurance mathematics. In: Engquist, B. and Schmid, W. (Eds) *Mathematics Unlimited, 2001 and Beyond*, pp. 703-722. Springer, Berlin.

[46] KLÜPPELBERG, C. AND PERGAMENCHTCHIKOV, S. (2004) The tail of the stationary distribution of a random coefficient AR(q) model. *Ann. Appl. Probab.* **14**, 971-1005.

[47] LEADBETTER, M.R., LINDGREN, G. AND ROOTZÉN, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

[48] LINDEBERG, Y.W. (1922) Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15**, 221-225.

[49] LINDVALL, T. (1992) *Lectures on the Coupling Method.* Wiley, New York.

[50] LONGIN, F.M. (2000) From value at risk to stress-testing: the extreme value approach. *J. Banking and Finance* **24**, 1097-1130.

[51] MEYN, S.P.. AND TWEEDIE, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer, London.

[52] MEYN, S.P.. AND TWEEDIE, R.L. (1993) Generalised resolvents and Harris recurrence of Markov processes. *Contemporary Mathematics* **149**, 227-250.

[53] MIKOSCH, T. AND STĂRICĂ, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427-1451.

[54] MOKKADEM, A. (1990) Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. Henri Poincaré* **26**, 219-260.

[55] NUMMELIN, E. AND TUOMINEN, P. (1982) Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **12**, 187-202.

[56] O'BRIEN, G. (1974) Limit theorems for the maximum term of a stationary process. *Ann. Probab.* **2**, 540-545.

[57] PICKANDS, J. (1981) Multivariate extreme value distributions. In: *43rd Sess. Int. Statist. Inst.*, 540-545.

[58] RACHEV, S.T. (1991) *Probability Metrics and the Stability of Stochastic Models.* Wiley, Chichester.

[59] RACHEV, S.T. AND RÜSCHENDORF, L. (1995) Probability metrics and recursive algorithms. *Adv. Appl. Probab.* **27**, 770-799.

[60] RACHEV, S.T. AND SAMORODNITSKY, G. (1995) Limit laws for a stochastic process and random recursion arising in probabilistic modelling. *Adv. Appl. Probab.* **27**, 185-202.

[61] RESNICK, S.I. (1987) *Extreme Values, Regular Variation and Point Processes.* Springer, New York.

[62] ROOS, B. (2003) Kerstan's method for compound Poisson approximation. *Ann. Probab.* **31**, 1754-1771.

[63] SILVERMAN, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

[64] STĂRICĂ, C (1999) Multivariate extremes for models with constant conditional correlations. *J. Empirical Finance* **6**, 515-553.

[65] VENABLES, W.N. AND RIPLEY, B.D. (1999) *Modern Applied Statistics with S-PLUS.* Springer, New York.

[66] VERETENNIKOV, A.Y. (1999) On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory Probab. Appl.* **44**, 361-374.

[67] VERVAAT, W. (1979) On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750-783.

[68] XIA, A. (1997) On the rate of Poisson process approximation to a Bernouilli process. *J. Appl. Probab.* **34**, 898-907.

List of symbols

$\mathbb{R}$	Set of the real numbers
$\mathbb{Z}$	Set of the integer numbers
$\mathbb{N}$	Set of the non-negative integer numbers
$\mathbb{N}^*$	$\mathbb{N}\setminus\{0\}$
$]a,b]$ with $(a,b)\in\mathbb{R}^2$	$\{x\in\mathbb{R}, a<x\leq b\}$
$[\![i,j[\![$ with $(i,j)\in\mathbb{Z}^2$	$\{k\in\mathbb{Z}, i\leq k<j\}$
$\lfloor x\rfloor$	$\max\{i\in\mathbb{N}, i\leq x\}$
$\lceil x\rceil$	$\min\{i\in\mathbb{N}, x\leq i\}$
x^+	$\max\{x,0\}$
S	Polish space
$\mathcal{S}$	σ-field on S
$\mathcal{B}(S)$	Borel σ-field on S
$(S,\mathcal{S})$	Measurable space, chosen as state space
Ω	Topological space
$\mathcal{A}$	σ-field on Ω

$\mathbb{P}$	Probability measure
$(\Omega, \mathcal{A}, \mathbb{P})$	Probability space
$\mathbb{E}$	Expectation
$\mathcal{L}(X)$	Distribution of X
F	Distribution function
$\overline{F}$	Tail of F, *i.e.* $1 - F$
i.i.d.	Independent and identically distributed
$\mathcal{M}(S)$	Set of the σ-finite measures on S
$\mathcal{N}(S)$	Set of the counting measures on S
$\mathcal{P}(S)$	Set of the point processes on S
$\mathcal{S}(S)$	Set of the measurable subsets of S
$\mathcal{F}(S)$	Set of the functionals of counting measures on S
$A \mapsto \delta_s(A)$	Dirac measure on $\mathcal{S}(S)$ at $s \in S$
$x \mapsto \mathbb{I}_{\{x \in A\}}$	Indicator function on S of $A \in \mathcal{S}(S)$
$\boldsymbol{\mu}(\bullet)$	Lebesgue measure
Π_n	Set of the permutations of $[\![1, n]\!]$
M_n	$\max\{X_1, \ldots, X_n\}$
$\widetilde{X}_1, \ldots, \widetilde{X}_n$	I.i.d. replicates of X_1
Poisson(λ)	Poisson law with intensity λ
CP(λ, ν)	Compound Poisson law
$X \stackrel{d}{=} Y$	X and Y have the same distribution
$X_n \underset{n\to+\infty}{\xrightarrow{d}} X$	$\mathcal{L}(X_n) \underset{n\to+\infty}{\longrightarrow} \mathcal{L}(X)$ weakly
$X_n \underset{n\to+\infty}{\xrightarrow{v}} X$	$\mathcal{L}(X_n) \underset{n\to+\infty}{\longrightarrow} \mathcal{L}(X)$ vaguely
$a_n \underset{n\to+\infty}{\sim} b_n$	$a_n/b_n \underset{n\to+\infty}{\longrightarrow} 1$
$a_n \ll b_n$ with $a_n, b_n \geq 0$	$a_n/b_n \underset{n\to+\infty}{\longrightarrow} 0$